FOR OFFICIAL USE ONLY 1916

U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office

SEARCH REQUEST FORM

Examiner # (**Mandatory**): _____ Requester's Full Name: _____

Art Unit _____ Location (Bldg/Room#): 10D11 Phone (circle 305 306 308) _____

Serial Number: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

Title of Invention _____

Inventors (please provide full names): _____

_____

Earliest Priority Date: _____

Keywords (include any known synonyms registry numbers, explanation of initialisms):

Search Topic:

Please write detailed statement of the search topic, and the concept of the invention. Describe as specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples of relevant citations, authors, etc., if known. You may include a copy of the abstract and the broadcast or most relevant claim(s).

STAFF USE ONLY

Searcher: Stephanie

Searcher Phone #: 305-4499

Searcher Location: _____

Date Picked Up: _____

Date Completed: 9/22/99

Clerical Prep Time: _____

Terminal Time: _____

Number of Databases: _____

Type of Search

_____ N.A. Sequence
_____ A.A. Sequence
_____ Structure (#)
_____ Bibliographic
_____ Litigation
_____ Fulltext
_____ Procurement
_____ Other

Vendors (include cost where applicable)

_____ STN
_____ Questel/Orbit
_____ Lexis/Nexis
_____ WWW/Internet
_____ In-house sequence systems (list)
_____ Dialog
_____ Dr. Link
_____ Westlaw
_____ Other (specify)

file copy

```
          *     8316     8332: gap of unknown length
          *     8333    10032: contig of 1700 bp in length
          *    10033    10049: gap of unknown length
          *    10050    11755: contig of 1706 bp in length
          *    11756    11772: gap of unknown length
          *    11773    20315: contig of 8543 bp in length
          *    20316    20332: gap of unknown length
          *    20333    38250: contig of 17918 bp in length
          *    38251    38267: gap of unknown length
          *    38268   115442: contig of 77175 bp in length
          *   115443   115459: gap of unknown length
          *   115460   198605: contig of 83146 bp in length.
FEATURES             Location/Qualifiers
     source          1. .198605
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DJ400N23"
BASE COUNT    53938 a  46957 c  46811 g  50746 t    153 others
ORIGIN

  Query Match           68.3%;  Score 1590;  DB 19;  Length 198605;
  Best Local Similarity  99.7%;  Pred. No. 0.00e+00;
  Matches  1605;  Conservative    0;  Mismatches   3;  Indels   2;  Gaps   2;

Db 125463 TATAGAGAAATGCCTTTATAGAAAAGTAGAAACCAGTAATATTCTCTTCTCCAGCATCAC 125522
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2316 TATAGAGAAATGCCTTTATAGAAAAGTAGAAACCAGTAATATTCTCTTCTCCAGCATCAC 2257

Db 125523 TAACACCAAGAGACCACCTGAGGTCTAGGTCCCCAAAGCAGATGGCTCCATAGAAAGCCC 125582
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2256 TAACACCAAGAGACCACCTGAGGTCTAGGTCCCCAAAGCAGATGGCTCCATAGAAAGCCC 2197

Db 125583 CACTAACCCCTGTCTCCACATTGGGCAGTGGAAGGGTTCTGGAAAGGAAGCTCTATGGCT 125642
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2196 CACTAACCCCTGTCTCCACATTGGGCAGTGGAAGGGTTCTGGAAAGGAAGCTCTATGGCT 2137

Db 125643 AGGAGCTGCCAAGGCCTCTTGAGTGTGACATCACAGGTTAGAGGCCCTGCTGAGCTGCTA 125702
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2136 AGGAGCTGCCAAGGCCTCTTGAGTGTGACATCACAGGTTAGAGGCCCTGCTGAGCTGCTA 2077

Db 125703 GCACAGTGCAGTGGAGAAATGTTGCGCTCCGTCAGGAAGCTGGGCTCTTGCCATGCCAGT 125762
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2076 GCACAGTGCAGTGGAGAAATGTTGCGCTCCGTCAGGAAGCTGGGCTCTTGCCATGCCAGT 2017

Db 125763 AGCTTGCTTACGACCTTGGGCAAGTCACTTTTTCCTCTCTGGGGTCCTGTGTCCAAAAGG 125822
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2016 AGCTTGCTTACGACCTTGGGCAAGTCACTTTTTCCTCTCTGGGGTCCTGTGTCCAAAAGG 1957

Db 125823 CTTAATGAGCCCTTTCTAATGGTTCTTTGATTGCTTTATGAGTTAGGGGCTGTGACCCTA 125882
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1956 CTTAATGAGCCCTTTCTAATGGTTCTTTGATTGCTTTATGAGTTAGGGGCTGTGACCCTA 1897

Db 125883 GCCCCAGTGCCCTACAGGAAGGTGCTGACCTGGACGCCCGCAGGACCACCTGCTAAATCA 125942
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Cp   1896 GCCCCAGTGCCCTACAGGAAGGTGCTGACCTGGACGCCCGCAG-ACCACCTGCTAAATCA 1838

Db 125943 GGCCCACTTGGGGAGCAAGCCCTGAAGTGCTTTATTACCTTGCCTCTGAGCTTAGTTTCC 126002
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1837 GGCCCACTTGGGGAGCAAGCCCTGAAGTGCTTTATTACCTTGCCTCTGAGCTTAGTTTCC 1778

Db 126003 AGGTTAGCCTCTTAGAAGCAGCTCACAAGGCCCCACAGTATAATTCCCATAGGTTAGCTA 126062
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1777 AGGTTAGCCTCTTAGAAGCAGCTCACAAGGCCCCACAGTATAATTCCCATAGGTTAGCTA 1718

Db 126063 ACTACTCTCTTACCTGTGCAAAATGAAGACAGGTCACTATCCATCCCTGAAGAGAAACTG 126122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1717 ACTACTCTCTTACCTGTGCAAAATGAAGACAGGTCACTATCCATCCCTGAAGAGAAACTG 1658

Db 126123 GGGCATCCCAGGTCTGCCTCTGCTCCTGTGGCCAAAGGCGGAGCAGATCCTTCTCTATCC 126182
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1657 GGGCATCCCAGGTCTGCCTCTGCTCCTGTGGCCAAAGGCGGAGCAGATCCTTCTCTATCC 1598

Db 126183 TCTTCCTAAGTGGAATGCTCAGTGCAATTCTTCCTTTCCTTATGAAACAGGAAGAGTCCC 126242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1597 TCTTCCTAAGTGGAATGCTCAGTGCAATTCTTCCTTTCCTTATGAAACAGGAAGAGTCCC 1538

Db 126243 TGGGCCCAGGCCTGGCCCACGGTTGTCAAGGCACATCATTGCCAGCAAGCTGAAGCATAC 126302
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1537 TGGGCCCAGGCCTGGCCCACGGTTGTCAAGGCACATCATTGCCAGCAAGCTGAAGCATAC 1478

Db 126303 CAGCAGCCACAACCTAGATCTCATTCCCAACCCAAAGTTCTGACTTCTGTACAAACTCGT 126362
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1477 CAGCAGCCACAACCTAGATCTCATTCCCAACCCAAAGTTCTGACTTCAGTACAAACTCGT 1418

Db 126363 TTCCAGTTCTCAGAGGTCAGTAGGGAGTTCCATCCCAAATATTTTAAAAACTAAATATAT 126422
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1417 TTCCAGTTGTCAGAGGTCAGTAGGGAGTTCCATCCCAAATATTTTAAAAACTAAATATAT 1358

Db 126423 TTTTTCAATGCCATCCCTTGACCTGCTTTACCTTAACCATGAATAACAGTGTATGCCACT 126482
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1357 TTTTTCAATGCCATCCCTTGACCTGCTTTACCTTAACCATGAATAACAGTGTATGCCACT 1298

Db 126483 CAAGCCCACAGTGGAGACTGAGGCCATCTTAGTAAAGCACAGAGGAGCCTGGGGGAGCTC 126542
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Cp   1297 CAAGCCCACAGTGGAGACTGAG-CCATCTTAGTAAAGCACAGAGGAGCCTGGGGGAGCTC 1239

Db 126543 CCCGTCTATCCCTGGTTCAGTTCCAGGGGTGCAGGCAATGTTTGGAAGCCCTTAGCAGCA 126602
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1238 CCCGTCTATCCCTGGTTCAGTTCCAGGGGTGCAGGCAATGTTTGGAAGCCCTTAGCAGCA 1179

Db 126603 GCATCACTGTGGGAGCTGCCACTGTCTCCATCCACAGACAAAATCTGCTCCAAGGGATGG 126662
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1178 GCATCACTGTGGGAGCTGCCACTGTCTCCATCCACAGACAAAATCTGCTCCAAGGGATGG 1119

Db 126663 ACAAGGAGCACTGTTAGGACCCTACCCAGCCTTACCCTCAGCCCACAGGGCCACCTGCTT 126722
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1118 ACAAGGAGCACTGTTAGGACCCTACCCAGCCTTACCCTCAGCCCACAGGGCCACCTGCTT 1059

Db 126723 CTGTCACTCTTACTAGGATTCGCCAAAAAAGCGGGGGAGTGGTAGGGTTTTAACCAGAAC 126782
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1058 CTGTCACTCTTACTAGGATTCGCCAAAAAAGCGGGGGAGTGGTAGGGTTTTAACCAGAAC 999

Db 126783 ACAAAGTGGTAGTTCCTCCTAGCTGTAGGAGGGTGGGCTGTCCTGCACCAGTGTCTGCAT 126842
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    998 ACAAAGTGGTAGTTCCTCCTAGCTGTAGGAGGGTGGGCTGTCCTGCACCAGTGTCTGCAT 939

Db 126843 GGGCTCCTGCCCACTGGGGGGGCTCAGGCCCCAGGAGTCCCGGCCTGGCCCATAAGGGGG 126902
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    938 GGGCTCCTGCCCACTGGGGGGGCTCAGGCCCCAGGAGTCCCGGCCTGGCCCATAAGGGGG 879

Db 126903 GTGGTGCCCTCCTGAGGGTCAACTGGAGTCTGGCTGGTGTGGACCACGACAGTCTTCTCA 126962
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    878 GTGCTGCCCTCCTGAGGGTCAACTGGAGTCTGGCTGGTGTGGACCACGACAGTCTTCTCA 819

Db 126963 TCCACAATCTCACAGGTGGGGTTGGTGAAGGCAGACAAGGGCAGAGTGATTCGCTGCATC 127022
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    818 TCCACAATCTCACAGGTGGGGTTGGTGAAGGCAGACAAGGGCAGAGTGATTCGCTGCATC 759

Db 127023 TCCCTCTCACATACTTTCTGATCATGCTGTTCTTTCAAATCCTTCCCCCT 127072
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    758 TCCCTCTCACATACTTTCTGATCATGCTGTTCTTTCAAATCCTTCCCCCT 709


RESULT    2
LOCUS       HSAC002073 128978 bp    DNA              PRI       12-MAY-1997
DEFINITION  Human PAC clone DJ515N1 from 22q11.2-q22, complete sequence.
ACCESSION   AC002073
NID         g2078469
VERSION     AC002073.1  GI:2078469
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 128978)
```

AUTHORS Du,Z, Scheet,P and Harper,M.
TITLE The sequence of H. sapiens PAC clone DJ515N1
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 128978)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (12-MAY-1997)
COMMENT SUBMITTED BY:
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre chromosome 22 mapping group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr22/

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-3 prepared by Pieter de Jong and coworkers at Roswell Park Cancer Institute, using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor. For further details, see http://bacpac.med.buffalo.edu/ The clone is available from Genome Systems, Inc. (http://www.genomesystems.com).
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is H_DJ400N23; the clone sequenced to the right is H_DJ412A9. Actual start of this clone is at base position 1 of H_DJ515N1.

This clone contains STS WI-12936 (NID:g1344756) and A006I21 (NID:g1341182).

FEATURES Location/Qualifiers

source	1. .128978
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="22"
	/clone="DJ515N1"
	/clone_lib="RPCI-3"
	/map="22q11.2-q22"
repeat_region	7. .131
	/rpt_family="L1"
repeat_region	223. .526
	/rpt_family="ALU"
repeat_region	528. .944
	/rpt_family="L1"
repeat_region	989. .1266
	/rpt_family="ALU"
repeat_region	1485. .1576
	/rpt_family="L1"
repeat_region	complement(1599. .1626)
	/rpt_family="MER"
repeat_region	1620. .1939
	/rpt_family="MER"
repeat_region	complement(1959. .2201)
	/rpt_family="MER"
repeat_region	complement(2813. .3101)
	/rpt_family="ALU"
repeat_region	complement(3383. .3639)
	/rpt_family="ALU"
repeat_region	5560. .5648
	/rpt_family="ALU"
repeat_region	complement(6859. .7148)
	/rpt_family="ALU"
repeat_region	8165. .8447
	/rpt_family="ALU"
repeat_region	complement(8453. .8742)
	/rpt_family="ALU"
repeat_region	9147. .9438
	/rpt_family="ALU"
repeat_region	11744. .12034
	/rpt_family="ALU"
repeat_region	complement(12037. .12313)
	/rpt_family="ALU"
repeat_region	12643. .12925
	/rpt_family="ALU"
repeat_region	13532. .13693
	/rpt_family="ALU"
repeat_region	13889. .14009
	/rpt_family="ALU"
misc_feature	14446. .14613
	/note="match to human EST AA122029 (NID:g1678048) zm25f10.r1"
misc_feature	14452. .14613
	/note="match to human EST D31562 (NID:g644442)"
gene	14543. .23803
	/gene="WUGSC:DJ515N1.2"
CDS	join(14543. .14612,15591. .15707,15803. .15922,17188. .17388, 17494. .17572,23599. .23803)
	/gene="WUGSC:DJ515N1.2"
	/note="Putative gene. Genscan predictions confirmed by EST splicing.; coded for by human cDNAs AA122029 (NID:g1678048), D31562 (NID:g644442), AA158721 (NID:g1733515), R59640 (NID:g830335) and F13082 (NID:g709111)"
	/codon_start=1
	/protein_id="AAB54054.1"
	/db_xref="PID:g2078470"
	/db_xref="GI:2078470"
	/translation="MLLAWVQAFLVSNMLLAEAYGSGGCFWDNGHLYREDQTSPAPGL RCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRGPWCYVSGEAGVPEKRPCEDLRCP ETTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARSEAAAVQPVIGISQRVRMNS KEKKDLGTLGYVLGITMMVIIIAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRITL PLSAFTNPTCEIVDEKTVVVHTSQTPVDPQEGTTPLMGQAGTPGA"
misc_feature	14544. .14613
	/gene="WUGSC:DJ515N1.2"
	/note="match to human EST AA158721 (NID:g1733515) zo78g11.r1"
misc_feature	15589. .15641
	/gene="WUGSC:DJ515N1.2"
	/note="match to human EST D31562 (NID:g644442)"
misc_feature	15589. .15708
	/gene="WUGSC:DJ515N1.2"
	/note="match to human EST AA158721 (NID:g1733515) zo78g11.r1"
misc_feature	15589. .15706
	/gene="WUGSC:DJ515N1.2"
	/note="match to human EST AA122029 (NID:g1678048) zm25f10.r1"
repeat_region	complement(16489. .16778)
	/rpt_family="ALU"
repeat_region	complement(18503. .18758)
	/rpt_family="ALU"
repeat_region	19258. .19548
	/rpt_family="ALU"
repeat_region	19569. .19858
	/rpt_family="ALU"
repeat_region	complement(20735. .20996)
	/rpt_family="ALU"

```
     repeat_region   22684. .22975
                     /rpt_family="ALU"
     repeat_region   23163. .23289
                     /rpt_family="ALU"
     misc_feature    23942. .24240
                     /note="match to human EST Z44818 (NID:g573984)"
     repeat_region   complement(25449. .25740)
                     /rpt_family="ALU"
     gene            complement(28446. .58170)
                     /gene="LIMK-2"
     gene            complement(28446. .94478)
                     /gene="LIMK-2"
     CDS             complement(join(28446. .28590,31579. .31736,33380. .33435,
                     34183. .34357,35686. .35751,38687. .38743,38980. .39111,
                     39817. .39903,40755. .40941,44096. .44292,44648. .44753,
                     46810. .46998,47657. .47766,48461. .48596,81068. .81167,
                     94463. .94478))
                     /gene="LIMK-2"
                     /note="DJ515N1.1b; match to D45906 (NID:g1136921); coded
                     for by human cDNA D45906 (NID:g1805593)"
                     /codon_start=1
                     /product="Lim Kinase"
                     /protein_id="AAB54056.1"
                     /db_xref="PID:g2078472"
                     /db_xref="GI:2078472"
                     /translation="MSALAGEDVWRCPGCGDHIAPSQIWYRTVNETWHGSCFRCSECQ
                     DSLTNWYYEKDGKLYCPKDYWGKFGEFCHGCSLLMTGPFMVAGEFKYHPECFACMSCK
                     VIIEDGDAYALVQHATLYCGKCHNEVVLAPMFERLSTESVQEQLPYSVTLISMPATTE
                     GRRGFSVSVESACSNYATTVQVKEVNRMHISPNNRNAIHPGDRILEINGTPVRTLRVE
                     EVEDAISQTSQTLQLLIEHDPVSQRLDQLRLEARLAPHMQNAGHPHALSTLDTKENLE
                     GTLRRRSLRRSNSISKSPGPSSPKEPLLFSRDISRSESLRCSSSYSQQIFRPCDLIHG
                     EVLGKGFFGQAIKVTHKATGKVMVMKELIRCDEETQKTFLTEVKVMRSLDHPNVLKFI
                     GVLYKDKKLNLLTEYIEGGTLKDFLRSMDPFPWQQKVRFAKGIASGMAYLHSMCIIHR
                     DLNSHNCLIKLDKTVVVADFGLSRLIVEERKRAPMEKATTKKRTLRKNDRKKRYTVVG
                     NPYWMAPEMLNGKSYDETVDIFSFGIVLCEIIGQVYADPDCLPRTLDFGLNVKLFWEK
                     FVPTDCPPAFFPLAAICCRLEPESRPAFSKLEDSFEALSLYLGELGIPLPAELEELDH
                     TVSMQYGLTRDSPP"
     CDS             complement(join(28446. .28590,29751. .30098,31579. .31736,
                     33380. .33435,34183. .34357,35686. .35751,38687. .38743,
                     38980. .39111,39817. .39903,40755. .40941,44096. .44292,
                     44648. .44753,46810. .46998,47657. .47766,48461. .48596,
                     58118. .58170))
...
Note: remainder of annotations omitted.

  Query Match           67.7%;  Score 1576;  DB 30;  Length 128978;
  Best Local Similarity 99.6%;  Pred. No. 0.00e+00;
  Matches  1603;  Conservative    0;  Mismatches   3;  Indels   4;  Gaps   3;

Db  23597 AGGGGGAAGGATTTGAAAGAACAGCATGATCAGAAAGTATGTGAGAGGGAGATGCAGCGA 23656
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    709 AGGGGGAAGGATTTGAAAGAACAGCATGATCAGAAAGTATGTGAGAGGGAGATGCAGCGA 768

Db  23657 ATCACTCTGCCCTTGTCTGCCTTCACCAACCCCACCTGTGAGATTGTGGATGAGAAGACT 23716
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    769 ATCACTCTGCCCTTGTCTGCCTTCACCAACCCCACCTGTGAGATTGTGGATGAGAAGACT 828

Db  23717 GTCGTGGTCCACACCAGCCAGACTCCAGTTGACCCTCAGGAGGGCACCACCCCCCTTATG 23776
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Qy    829 GTCGTGGTCCACACCAGCCAGACTCCAGTTGACCCTCAGGAGGGCAGCACCCCCCTTATG 888

Db  23777 GGCCAGGCCGGGACTCCTGGGGCCTGAGCCCCCCCAGTGGGCAGGAGCCCATGCAGACAC 23836
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    889 GGCCAGGCCGGGACTCCTGGGGCCTGAGCCCCCCCAGTGGGCAGGAGCCCATGCAGACAC 948

Db  23837 TGGTGCAGGACAGCCCACCCTCCTACAGCTAGGAGGAACTACCACTTTGTGTTCTGGTTA 23896
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    949 TGGTGCAGGACAGCCCACCCTCCTACAGCTAGGAGGAACTACCACTTTGTGTTCTGGTTA 1008

Db  23897 AAACCCTACCACTCCCCCGCTTTTTTGGCGAATCCTAGTAAGAGTGACAGAAGCAGGTGG 23956
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1009 AAACCCTACCACTCCCCCGCTTTTTTGGCGAATCCTAGTAAGAGTGACAGAAGCAGGTGG 1068

Db  23957 CCCTGTGGGCTGAGGGTAAGGCTGGGTAGGGTCCTAACAGTGCTCCTTGTCCATCCCTTG 24016
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1069 CCCTGTGGGCTGAGGGTAAGGCTGGGTAGGGTCCTAACAGTGCTCCTTGTCCATCCCTTG 1128

Db  24017 GAGCAGATTTTGTCTGTGGATGGAGACAGTGGCAGCTCCCACAGTGATGCTGCTGCTAAG 24076
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1129 GAGCAGATTTTGTCTGTGGATGGAGACAGTGGCAGCTCCCACAGTGATGCTGCTGCTAAG 1188

Db  24077 GGCTTCCAAACATTGCCTGCACCCCTGGAACTGAACCAGGGATAGACGGGGAGCTCCCCC 24136
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1189 GGCTTCCAAACATTGCCTGCACCCCTGGAACTGAACCAGGGATAGACGGGGAGCTCCCCC 1248

Db  24137 AGGCTCCTCTGTGCTTTACTAAGATGGCCTCAGTCTCCACTGTGGGCTTGAGTGGCATAC 24196
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Qy   1249 AGGCTCCTCTGTGCTTTACTAAGATGGC-TCAGTCTCCACTGTGGGCTTGAGTGGCATAC 1307

Db  24197 ACTGTTATTCATGGTTAAGGTAAAGCAGGTCAAGGGATGGCATTGAAAAAATATATTTAG 24256
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1308 ACTGTTATTCATGGTTAAGGTAAAGCAGGTCAAGGGATGGCATTGAAAAAATATATTTAG 1367

Db  24257 TTTTTAAAATATTTGGGATGGAACTCCCTACTGACCTCTGAGAACTGGAAACGAGTTTGT 24316
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Qy   1368 TTTTTAAAATATTTGGGATGGAACTCCCTACTGACCTCTGACAACTGGAAACGAGTTTGT 1427

Db  24317 ACAGAAGTCAGAACTTTGGGTTGGGAATGAGATCTAGGTTGTGGCTGCTGGTATGCTTCA 24376
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1428 ACTGAAGTCAGAACTTTGGGTTGGGAATGAGATCTAGGTTGTGGCTGCTGGTATGCTTCA 1487

Db  24377 GCTTGCTGGCAATGATGTGCCTTGACAACCGTGGGCCAGGCCTGGGCCCAGGGACTCTTC 24436
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1488 GCTTGCTGGCAATGATGTGCCTTGACAACCGTGGGCCAGGCCTGGGCCCAGGGACTCTTC 1547

Db  24437 CTGTTTCATAAGGAAAGGAAGAATTGCACTGAGCATTCCACTTAGGAAGAGGATAGAGAA 24496
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1548 CTGTTTCATAAGGAAAGGAAGAATTGCACTGAGCATTCCACTTAGGAAGAGGATAGAGAA 1607

Db  24497 GGATCTGCTCCGCCTTTGGCCACAGGAGCAGAGGCAGACCTGGGATGCCCCAGTTTCTCT 24556
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1608 GGATCTGCTCCGCCTTTGGCCACAGGAGCAGAGGCAGACCTGGGATGCCCCAGTTTCTCT 1667

Db  24557 TCAGGGATGGATAGTGACCTGTCTTCATTTTGCACAGGTAAGAGAGTAGTTAGCTAACCT 24616
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1668 TCAGGGATGGATAGTGACCTGTCTTCATTTTGCACAGGTAAGAGAGTAGTTAGCTAACCT 1727

Db  24617 ATGGGAATTATACTGTGGGGCCTTGTGAGCTGCTTCTAAGAGGCTAACCTGGAAACTAAG 24676
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1728 ATGGGAATTATACTGTGGGGCCTTGTGAGCTGCTTCTAAGAGGCTAACCTGGAAACTAAG 1787

Db  24677 CTCAGAGGCAAGGTAATAAAGCACTTCAGGGCTTGCTCCCCAAGTGGGCCTGATTTAGCA 24736
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1788 CTCAGAGGCAAGGTAATAAAGCACTTCAGGGCTTGCTCCCCAAGTGGGCCTGATTTAGCA 1847

Db  24737 GGTGGTCCTGCGGGCGTCCAGGTCAGCACCTTCCTGTAGGGCACTGGGGCTAGGGTCACA 24796
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1848 GGTGGTC-TGCGGGCGTCCAGGTCAGCACCTTCCTGTAGGGCACTGGGGCTAGGGTCACA 1906

Db  24797 GCCCCTAACTCATAAAGCAATCAAAGAACCATTAGAAAGGGCTCATTAAGCCTTTTGGAC 24856
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1907 GCCCCTAACTCATAAAGCAATCAAAGAACCATTAGAAAGGGCTCATTAAGCCTTTTGGAC 1966

Db  24857 ACAGGACCCCAGAGAGGAAAAAGTGACTTGCCCAAGGTCGTAAGCAAGCTACTGGCATGG 24916
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1967 ACAGGACCCCAGAGAGGAAAAAGTGACTTGCCCAAGGTCGTAAGCAAGCTACTGGCATGG 2026

Db  24917 CAAGAGCCCAGCTTCCTGACGGAGCGCAACATTTCTCCACTGCACTGTGCTAGCAGCTCA 24976
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2027 CAAGAGCCCAGCTTCCTGACGGAGCGCAACATTTCTCCACTGCACTGTGCTAGCAGCTCA 2086

Db  24977 GCAGGGCCTCTAACCTGTGATGTCACACTCAAGAGGCCTTGGCAGCTCCTAGCCATAGAG 25036
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2087 GCAGGGCCTCTAACCTGTGATGTCACACTCAAGAGGCCTTGGCAGCTCCTAGCCATAGAG 2146

Db  25037 CTTCCTTTCCAGAACCCTTCCACTGCCCAATGTGGAGAC--GGGTTAGTGGGGCTTTCTA 25094
```

```
Qy   2147 CTTCCTTTCCAGAACCCTTCCACTGCCCAATGTGGAGACAGGGGTTAGTGGGGCTTTCTA 2206

Db  25095 TGGAGCCATCTGCTTTGGGGACCTAGACCTCAGGTGGTCTCTTGGTGTTAGTGATGCTGG 25154
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2207 TGGAGCCATCTGCTTTGGGGACCTAGACCTCAGGTGGTCTCTTGGTGTTAGTGATGCTGG 2266

Db  25155 AGAAGAGAATATTACTGGTTTCTACTTTTCTATAAAGGCATTTCTCTATA 25204
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2267 AGAAGAGAATATTACTGGTTTCTACTTTTCTATAAAGGCATTTCTCTATA 2316


RESULT    3
LOCUS       AC005478    150140 bp    DNA              HTG       14-AUG-1998
DEFINITION  Homo sapiens clone DJ0425I02, WORKING DRAFT SEQUENCE, 5 unordered
            pieces.
ACCESSION   AC005478
NID         g3419864
VERSION     AC005478.1  GI:3419864
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150140)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 150140)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1987: contig of 1987 bp in length
            *     1988     2005: gap of unknown length
            *     2006     4759: contig of 2754 bp in length
            *     4760     4777: gap of unknown length
            *     4778    12352: contig of 7575 bp in length
            *    12353    12370: gap of unknown length
            *    12371    46730: contig of 34360 bp in length
            *    46731    46748: gap of unknown length
            *    46749   150140: contig of 103392 bp in length.
FEATURES             Location/Qualifiers
     source          1. .150140
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DJ0425I02"
BASE COUNT    38079 a  35586 c  35793 g  40610 t     72 others
ORIGIN

  Query Match           67.7%;  Score 1576;  DB 19;  Length 150140;
  Best Local Similarity  99.6%;  Pred. No. 0.00e+00;
  Matches  1603;  Conservative    0;  Mismatches   3;  Indels   4;  Gaps   3;

Db  76991 TATAGAGAAATGCCTTTATAGAAAAGTAGAAACCAGTAATATTCTCTTCTCCAGCATCAC 77050
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2316 TATAGAGAAATGCCTTTATAGAAAAGTAGAAACCAGTAATATTCTCTTCTCCAGCATCAC 2257

Db  77051 TAACACCAAGAGACCACCTGAGGTCTAGGTCCCCAAAGCAGATGGCTCCATAGAAAGCCC 77110
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2256 TAACACCAAGAGACCACCTGAGGTCTAGGTCCCCAAAGCAGATGGCTCCATAGAAAGCCC 2197

Db  77111 CACTAACCC--GTCTCCACATTGGGCAGTGGAAGGGTTCTGGAAAGGAAGCTCTATGGCT 77168
          |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2196 CACTAACCCCTGTCTCCACATTGGGCAGTGGAAGGGTTCTGGAAAGGAAGCTCTATGGCT 2137

Db  77169 AGGAGCTGCCAAGGCCTCTTGAGTGTGACATCACAGGTTAGAGGCCCTGCTGAGCTGCTA 77228
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2136 AGGAGCTGCCAAGGCCTCTTGAGTGTGACATCACAGGTTAGAGGCCCTGCTGAGCTGCTA 2077

Db  77229 GCACAGTGCAGTGGAGAAATGTTGCGCTCCGTCAGGAAGCTGGGCTCTTGCCATGCCAGT 77288
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2076 GCACAGTGCAGTGGAGAAATGTTGCGCTCCGTCAGGAAGCTGGGCTCTTGCCATGCCAGT 2017

Db  77289 AGCTTGCTTACGACCTTGGGCAAGTCACTTTTTCCTCTCTGGGGTCCTGTGTCCAAAAGG 77348
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2016 AGCTTGCTTACGACCTTGGGCAAGTCACTTTTTCCTCTCTGGGGTCCTGTGTCCAAAAGG 1957

Db  77349 CTTAATGAGCCCTTTCTAATGGTTCTTTGATTGCTTTATGAGTTAGGGGCTGTGACCCTA 77408
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1956 CTTAATGAGCCCTTTCTAATGGTTCTTTGATTGCTTTATGAGTTAGGGGCTGTGACCCTA 1897

Db  77409 GCCCCAGTGCCCTACAGGAAGGTGCTGACCTGGACGCCCGCAGGACCACCTGCTAAATCA 77468
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Cp   1896 GCCCCAGTGCCCTACAGGAAGGTGCTGACCTGGACGCCCGCAG-ACCACCTGCTAAATCA 1838

Db  77469 GGCCCACTTGGGGAGCAAGCCCTGAAGTGCTTTATTACCTTGCCTCTGAGCTTAGTTTCC 77528
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1837 GGCCCACTTGGGGAGCAAGCCCTGAAGTGCTTTATTACCTTGCCTCTGAGCTTAGTTTCC 1778

Db  77529 AGGTTAGCCTCTTAGAAGCAGCTCACAAGGCCCCACAGTATAATTCCCATAGGTTAGCTA 77588
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1777 AGGTTAGCCTCTTAGAAGCAGCTCACAAGGCCCCACAGTATAATTCCCATAGGTTAGCTA 1718

Db  77589 ACTACTCTCTTACCTGTGCAAAATGAAGACAGGTCACTATCCATCCCTGAAGAGAAACTG 77648
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1717 ACTACTCTCTTACCTGTGCAAAATGAAGACAGGTCACTATCCATCCCTGAAGAGAAACTG 1658

Db  77649 GGGCATCCCAGGTCTGCCTCTGCTCCTGTGGCCAAAGGCGGAGCAGATCCTTCTCTATCC 77708
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1657 GGGCATCCCAGGTCTGCCTCTGCTCCTGTGGCCAAAGGCGGAGCAGATCCTTCTCTATCC 1598

Db  77709 TCTTCCTAAGTGGAATGCTCAGTGCAATTCTTCCTTTCCTTATGAAACAGGAAGAGTCCC 77768
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1597 TCTTCCTAAGTGGAATGCTCAGTGCAATTCTTCCTTTCCTTATGAAACAGGAAGAGTCCC 1538

Db  77769 TGGGCCCAGGCCTGGCCCACGGTTGTCAAGGCACATCATTGCCAGCAAGCTGAAGCATAC 77828
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1537 TGGGCCCAGGCCTGGCCCACGGTTGTCAAGGCACATCATTGCCAGCAAGCTGAAGCATAC 1478

Db  77829 CAGCAGCCACAACCTAGATCTCATTCCCAACCCAAAGTTCTGACTTCTGTACAAACTCGT 77888
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Cp   1477 CAGCAGCCACAACCTAGATCTCATTCCCAACCCAAAGTTCTGACTTCAGTACAAACTCGT 1418

Db  77889 TTCCAGTTCTCAGAGGTCAGTAGGGAGTTCCATCCCAAATATTTTAAAAACTAAATATAT 77948
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1417 TTCCAGTTGTCAGAGGTCAGTAGGGAGTTCCATCCCAAATATTTTAAAAACTAAATATAT 1358

Db  77949 TTTTTCAATGCCATCCCTTGACCTGCTTTACCTTAACCATGAATAACAGTGTATGCCACT 78008
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1357 TTTTTCAATGCCATCCCTTGACCTGCTTTACCTTAACCATGAATAACAGTGTATGCCACT 1298

Db  78009 CAAGCCCACAGTGGAGACTGAGGCCATCTTAGTAAAGCACAGAGGAGCCTGGGGGAGCTC 78068
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Cp   1297 CAAGCCCACAGTGGAGACTGAG-CCATCTTAGTAAAGCACAGAGGAGCCTGGGGGAGCTC 1239

Db  78069 CCCGTCTATCCCTGGTTCAGTTCCAGGGGTGCAGGCAATGTTTGGAAGCCCTTAGCAGCA 78128
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1238 CCCGTCTATCCCTGGTTCAGTTCCAGGGGTGCAGGCAATGTTTGGAAGCCCTTAGCAGCA 1179

Db  78129 GCATCACTGTGGGAGCTGCCACTGTCTCCATCCACAGACAAAATCTGCTCCAAGGGATGG 78188
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1178 GCATCACTGTGGGAGCTGCCACTGTCTCCATCCACAGACAAAATCTGCTCCAAGGGATGG 1119

Db  78189 ACAAGGAGCACTGTTAGGACCCTACCCAGCCTTACCCTCAGCCCACAGGGCCACCTGCTT 78248
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1118 ACAAGGAGCACTGTTAGGACCCTACCCAGCCTTACCCTCAGCCCACAGGGCCACCTGCTT 1059
```

```
Db  78249 CTGTCACTCTTACTAGGATTCGCCAAAAAAGCGGGGGAGTGGTAGGGTTTTAACCAGAAC 78308
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1058 CTGTCACTCTTACTAGGATTCGCCAAAAAAGCGGGGGAGTGGTAGGGTTTTAACCAGAAC 999

Db  78309 ACAAAGTGGTAGTTCCTCCTAGCTGTAGGAGGGTGGGCTGTCCTGCACCAGTGTCTGCAT 78368
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    998 ACAAAGTGGTAGTTCCTCCTAGCTGTAGGAGGGTGGGCTGTCCTGCACCAGTGTCTGCAT 939

Db  78369 GGGCTCCTGCCCACTGGGGGGGCTCAGGCCCCAGGAGTCCCGGCCTGGCCCATAAGGGGG 78428
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    938 GGGCTCCTGCCCACTGGGGGGGCTCAGGCCCCAGGAGTCCCGGCCTGGCCCATAAGGGGG 879

Db  78429 GTGGTGCCCTCCTGAGGGTCAACTGGAGTCTGGCTGGTGTGGACCACGACAGTCTTCTCA 78488
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    878 GTGCTGCCCTCCTGAGGGTCAACTGGAGTCTGGCTGGTGTGGACCACGACAGTCTTCTCA 819

Db  78489 TCCACAATCTCACAGGTGGGGTTGGTGAAGGCAGACAAGGGCAGAGTGATTCGCTGCATC 78548
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    818 TCCACAATCTCACAGGTGGGGTTGGTGAAGGCAGACAAGGGCAGAGTGATTCGCTGCATC 759

Db  78549 TCCCTCTCACATACTTTCTGATCATGCTGTTCTTTCAAATCCTTCCCCCT 78598
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    758 TCCCTCTCACATACTTTCTGATCATGCTGTTCTTTCAAATCCTTCCCCCT 709
```

```
RESULT    4
LOCUS       G24591         464 bp    DNA              STS       31-MAY-1996
DEFINITION  human STS WI-14622.
ACCESSION   G24591
NID         g1344917
VERSION     G24591.1  GI:1344917
KEYWORDS    STS sequence; primer; sequence tagged site.
SOURCE      human STSs derived from sequences in dbEST and the Unigene
            collection.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped STSs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu

            Primer A: CCAGTAATATTCTCTTCTCCAGCA
            Primer B: ACCCTTCCACTGCCCAAT
            STS size: 125
            PCR Profile:
                  Presoak:
                  Denaturation:
                  Annealing: 56 degrees C
                  Polymerization:
                  PCR Cycles: 35
                  Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
                Taq Polymerase: 0.025 units/ul
                Total Vol: 20 ul

            Buffer:
                MgCl2: 1.5 mM
                KCl: 50 mM
                Tris-HCL: 10 mM
                pH: 9.3

            Derived from dbEST (genbank accession R59582).
FEATURES             Location/Qualifiers
     source          1. .464
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="99.7 cR from top of Chr22 linkage group"
     STS             43. .167
     primer_bind     43. .66
     primer_bind     complement(150. .167)
BASE COUNT      109 a    108 c    119 g    125 t      3 others
ORIGIN
```

```
  Query Match           17.3%;  Score 403;  DB 34;  Length 464;
  Best Local Similarity  96.5%;  Pred. No. 0.00e+00;
  Matches   447;  Conservative    0;  Mismatches  10;  Indels   6;  Gaps   6;

Db      1 TTTTTTTTTTTATAGAGAAATGCCTTTATAGAAAAGTAGAAACCAGTAATATTCTCTTCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2326 TTTTTTTTTTTATAGAGAAATGCCTTTATAGAAAAGTAGAAACCAGTAATATTCTCTTCT 2267

Db     61 CCAGCATCACTAACACCAAGAGACCACCTGAGGTCTAGGTCCCCAAAGCAGATGGCTCCA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2266 CCAGCATCACTAACACCAAGAGACCACCTGAGGTCTAGGTCCCCAAAGCAGATGGCTCCA 2207

Db    121 TAGAAAGCCCCACTAACCCCTGTCTCCACATTGGGCAGTGGAAGGGTTCTGGAAAGGAAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2206 TAGAAAGCCCCACTAACCCCTGTCTCCACATTGGGCAGTGGAAGGGTTCTGGAAAGGAAG 2147

Db    181 CTCTATGGCTAGGAGCTGCCAAGGCCTCTTGAGTGTGACATCACAGGTTAGAGGCCCTGC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2146 CTCTATGGCTAGGAGCTGCCAAGGCCTCTTGAGTGTGACATCACAGGTTAGAGGCCCTGC 2087

Db    241 TGAGCTGCTAGCACAGTGCAGTGGAGAAATGTTGCGCTCCGTCAG-AAGCTGGGCTCTTG 299
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Cp   2086 TGAGCTGCTAGCACAGTGCAGTGGAGAAATGTTGCGCTCCGTCAGGAAGCTGGGCTCTTG 2027

Db    300 CCATGCCAGTAGCTTGCTTACGACCTTGGGCAAGTCACTTTTTCCTCTCTGGGGGTCCTG 359
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Cp   2026 CCATGCCAGTAGCTTGCTTACGACCTTGGGCAAGTCACTTTTTCCTCTCTGGGG-TCCTG 1968

Db    360 TGTTCCAAAAGGTTTAATGAGGCCCTTTNTTAATGGTTCTTTGATTGNTTATGGAGTTAG 419
          ||| |||||||| |||||||| |||||| | ||||||||||||||||| ||   |||||||
Cp   1967 TGT-CCAAAAGGCTTAATGAG-CCCTTTCT-AATGGTTCTTTGATTGCTTTATGAGTTAG 1911

Db    420 GGGTT-TGACCTAGGCCCCAGTGCCCTACAGGAAGGTGCTGAC 461
          ||| | |||||   ||||||||||||||||||||||||||||||
Cp   1910 GGGCTGTGACCCTAGCCCCAGTGCCCTACAGGAAGGTGCTGAC 1868
```

```
RESULT    5
LOCUS       G20845         239 bp    DNA              STS       24-JUL-1996
DEFINITION  human STS A006I21, sequence tagged site.
ACCESSION   G20845
NID         g1341182
VERSION     G20845.1  GI:1341182
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 239)
  AUTHORS   Adams,M.D.
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Mark Adams
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Email: mdadams@tigr.org

            Primer A: GTCAGAATGAGGTATATAAAACA
            Primer B: GTGATGTCACACTCAAGA
```

STS size: 239
PCR Profile:
Denaturation: 96C 5min
Anneal: 54C 30sec
Extend: 72C 30sec
Denature: 95C 30sec
FinalExtend: 72C 5min
Cycles: 30
Protocol:
GenomicDNA: 25 ng
Primer: 0.43 uM each
dNTPs: 230 uM each
AmpliTaq: 0.5 units
TaqStart Ab: 0.5 units
Total Volume: 10 ul

Buffer:
Tris-HCl pH8.8: 100 mM
KCl: 500 mM
MgCl2: 20 mM
Triton X-100: 1%
Concentration: 10X

Prepared with primer pairs derived from THC85933: GenBank Accession Numbers- H27162, H27370, R59582, F03785, F10676, Z40604, N31943, N42779.

FEATURES Location/Qualifiers
source 1. .239
/organism="Homo sapiens"
/db_xref="taxon:9606"
STS 1. .239
primer_bind 1. .23
primer_bind complement(222. .239)
BASE COUNT 77 a 56 c 54 g 52 t
ORIGIN

Query Match 8.6%; Score 200; DB 34; Length 239;
Best Local Similarity 99.1%; Pred. No. 1.56e-137;
Matches 212; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

```
Db      28 TATAGAGAAATGCCTTTATAGAAAAGTAGAAACCAGTAATATTCTCTTCTCCAGCATCAC 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    2316 TATAGAGAAATGCCTTTATAGAAAAGTAGAAACCAGTAATATTCTCTTCTCCAGCATCAC 2257

Db      88 TAACACCAAGAGACCACCTGAGGTCTAGGTCCCCAAAGCAGATGGCTCCATAGAAAGCCC 147
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    2256 TAACACCAAGAGACCACCTGAGGTCTAGGTCCCCAAAGCAGATGGCTCCATAGAAAGCCC 2197

Db     148 CACTAACCC--GTCTCCACATTGGGCAGTGGAAGGGTTCTGGAAAGGAAGCTCTATGGCT 205
           |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Cp    2196 CACTAACCCCTGTCTCCACATTGGGCAGTGGAAGGGTTCTGGAAAGGAAGCTCTATGGCT 2137

Db     206 AGGAGCTGCCAAGGCCTCTTGAGTGTGACATCAC 239
           |||||||||||||||||||||||||||||||||
Cp    2136 AGGAGCTGCCAAGGCCTCTTGAGTGTGACATCAC 2103
```

RESULT 6
LOCUS G24430 284 bp DNA STS 31-MAY-1996
DEFINITION human STS WI-12936.
ACCESSION G24430
NID g1344756
VERSION G24430.1 GI:1344756
KEYWORDS STS sequence; primer; sequence tagged site.
SOURCE human STSs derived from sequences in dbEST and the Unigene collection.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 284)
AUTHORS Hudson,T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically Mapped STSs
JOURNAL Unpublished (1995)
COMMENT Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: CTATATGCAGGTGTCAGAATGAGG
Primer B: TTTCCAGAACCCTTCCACTG
STS size: 204
PCR Profile:
Presoak:
Denaturation:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:
Protocol:
Template: 10 ng
Primer: each 5 pM
dNTPs: each 4 nM
Taq Polymerase: 0.025 units/ul
Total Vol: 20 ul

Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCL: 10 mM
pH: 9.3

Derived from dbEST (genbank accession Z40604).

FEATURES Location/Qualifiers
source 1. .284
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="101.4 cR from top of Chr22 linkage group"
STS 49. .252
primer_bind 49. .72
primer_bind complement(233. .252)
BASE COUNT 90 a 61 c 59 g 71 t 3 others
ORIGIN

Query Match 8.3%; Score 194; DB 34; Length 284;
Best Local Similarity 98.5%; Pred. No. 2.16e-132;
Matches 194; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Db      88 TATAGAGAAATGCCTTTATAGAAAAGTAGAAACCAGTAATATTCTNTNCTCCAGCATCAC 147
           |||||||||||||||||||||||||||||||||||||||||||||| | ||||||||||||
Cp    2316 TATAGAGAAATGCCTTTATAGAAAAGTAGAAACCAGTAATATTCTCTTCTCCAGCATCAC 2257

Db     148 TAACACCAAGAGACCACCTGAGGTCTAGGTCCCCAAAGCAGATGGCTCCATAGAAAGCCC 207
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    2256 TAACACCAAGAGACCACCTGAGGTCTAGGTCCCCAAAGCAGATGGCTCCATAGAAAGCCC 2197

Db     208 CACTAACCCCTGTNTCCACATTGGGCAGTGGAAGGGTTCTGGAAAGGAAGCTCTATGGCT 267
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Cp    2196 CACTAACCCCTGTCTCCACATTGGGCAGTGGAAGGGTTCTGGAAAGGAAGCTCTATGGCT 2137

Db     268 AGGAGCTGCCAAGGCCT 284
           |||||||||||||||||
Cp    2136 AGGAGCTGCCAAGGCCT 2120
```

RESULT 7
LOCUS I66494 7218 bp DNA PAT 23-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
NID g2724471
VERSION I66494.1 GI:2724471
KEYWORDS .

provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="GDB:5740622"
/db_xref="taxon:9606"
/clone="IMAGE:701186"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"

BASE COUNT 145 a 130 c 148 g 120 t
ORIGIN

Query Match 22.2%; Score 518; DB 10; Length 543;
Best Local Similarity 99.3%; Pred. No. 0.00e+00;
Matches 542; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

```
Db      1 CTTCCTGTTTCATAAGGAAAGGAAGAATTGCACTGAGCATTCCACTTAGGAAGAGGATAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1544 CTTCCTGTTTCATAAGGAAAGGAAGAATTGCACTGAGCATTCCACTTAGGAAGAGGATAG 1603

Db     61 AGAAGGATCTGCTCCGCCTTTGGC-ACAGGAGCAGAGGCAGACCTGGGATGCCCCAGTTT 119
          ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Qy   1604 AGAAGGATCTGCTCCGCCTTTGGCCACAGGAGCAGAGGCAGACCTGGGATGCCCCAGTTT 1663

Db    120 CTCTTCAGGGATGGATAGTGACCTGTCTTCATTTTGCACAGGTAAGAGAGTAGTTAGCTA 179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1664 CTCTTCAGGGATGGATAGTGACCTGTCTTCATTTTGCACAGGTAAGAGAGTAGTTAGCTA 1723

       180 ACCTATGGGAATTATACTGTGGGGCCTTGTGAGCTGCTTCTAAGAGGCTAACCTGGAAAC 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1724 ACCTATGGGAATTATACTGTGGGGCCTTGTGAGCTGCTTCTAAGAGGCTAACCTGGAAAC 1783

Db    240 TAAGCTCAGAGGCAAGGTAATAAAGCACTTCAGGGCTTGCTCCCCAAGTGGGCCTGATTT 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1784 TAAGCTCAGAGGCAAGGTAATAAAGCACTTCAGGGCTTGCTCCCCAAGTGGGCCTGATTT 1843

Db    300 AGCAGGTGGTCCTGCGGGCGTCCAGGTCAGCACCTTCCTGTAGGGCACTGGGGCTAGGGT 359
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1844 AGCAGGTGGTC-TGCGGGCGTCCAGGTCAGCACCTTCCTGTAGGGCACTGGGGCTAGGGT 1902

Db    360 CACAGCCCCTAACTCATAAAGCAATCAAAGAACCATTAGAAAGGGCTCATTAA-CCTTTT 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Qy   1903 CACAGCCCCTAACTCATAAAGCAATCAAAGAACCATTAGAAAGGGCTCATTAAGCCTTTT 1962

Db    419 GGACACAGGACCCCAGAGAGGAAAAAGTGACTTGCCCAAGGTCGTAAGCAAGCTACTGGC 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1963 GGACACAGGACCCCAGAGAGGAAAAAGTGACTTGCCCAAGGTCGTAAGCAAGCTACTGGC 2022

Db    479 ATGGCAAGAGCC-AGCTTCCTGACGGAGCGCAACATTTCTCCACTGCACTGTGCTAGCAG 537
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2023 ATGGCAAGAGCCCAGCTTCCTGACGGAGCGCAACATTTCTCCACTGCACTGTGCTAGCAG 2082

Db    538 CTCAGC 543
          ||||||
Qy   2083 CTCAGC 2088
```

RESULT 2
LOCUS N29083 483 bp mRNA EST 04-JAN-1996
DEFINITION yx41f01.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone IMAGE:264313 5', mRNA sequence.
ACCESSION N29083
NID g1147319
VERSION N29083.1 GI:1147319
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 483)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On May 9, 1995 this sequence version replaced gi:802409.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 452
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1501 Std Error: 0.00
Seq primer: T7
High quality sequence stop: 452.

FEATURES Location/Qualifiers
source 1. .483
/organism="Homo sapiens"
/note="Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCAGTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P. Albino."
/db_xref="GDB:3873955"
/db_xref="taxon:9606"
/map="MMU16C3-C4 region"
/clone="IMAGE:264313"
/clone_lib="Soares melanocyte 2NbHM"
/sex="Male"
/tissue_type="melanocyte"
/lab_host="DH10B (ampicillin resistant)"

BASE COUNT 108 a 105 c 139 g 130 t 1 others
ORIGIN

Query Match 20.2%; Score 471; DB 32; Length 483;
Best Local Similarity 99.4%; Pred. No. 0.00e+00;
Matches 479; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

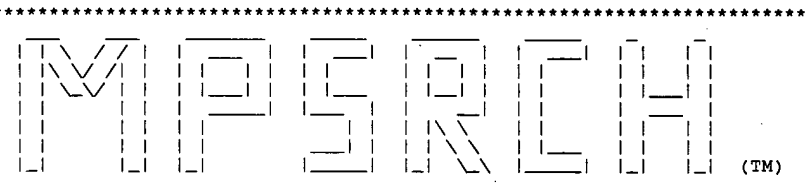
```
Db      1 GGCTGAGGGTAAGGCTGGGTAGGGTCCTAACAGTGCTCCTTGTCCATCCCTTGGAGCAGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1076 GGCTGAGGGTAAGGCTGGGTAGGGTCCTAACAGTGCTCCTTGTCCATCCCTTGGAGCAGA 1135

Db     61 TTTTGTCTGTGGATGGAGACAGTGGCAGCTCCCACAGTGATGCTGCTGCTAAGGGCTTCC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1136 TTTTGTCTGTGGATGGAGACAGTGGCAGCTCCCACAGTGATGCTGCTGCTAAGGGCTTCC 1195

Db    121 AAACATTGCCTGCACCCCTGGAACTGAACCAGGGATAGACGGGGAGCTCCCCCAGGCTCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1196 AAACATTGCCTGCACCCCTGGAACTGAACCAGGGATAGACGGGGAGCTCCCCCAGGCTCC 1255

Db    181 TCTGTGCTTTACTAAGATGGCCTCAGTCTCCACTGTGGGCTTGAGTGGCATACACTGTTA 240
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Qy   1256 TCTGTGCTTTACTAAGATGGC-TCAGTCTCCACTGTGGGCTTGAGTGGCATACACTGTTA 1314

Db    241 TTCATGGTTAAGGTAAAGCAGGTCAAGGGATGGCATTGAAAAAATATATTTAGTTTTTAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1315 TTCATGGTTAAGGTAAAGCAGGTCAAGGGATGGCATTGAAAAAATATATTTAGTTTTTAA 1374
```

```
Db     301 AATATTTGGGATGGAACTCCCTACTGACCTCTGAGAACTGGAAACGAGTTTGTACAGAAG 360
           ||||||||||||||||||||||||||||||||| |||||||||||||||||||| ||||
Qy    1375 AATATTTGGGATGGAACTCCCTACTGACCTCTGACAACTGGAAACGAGTTTGTACTGAAG 1434

Db     361 TCAGAACTTTGGGTTGGGAATGAGATCTAGGTTGTGGCTGCTGGTATGCTTCAGCTTGCT 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1435 TCAGAACTTTGGGTTGGGAATGAGATCTAGGTTGTGGCTGCTGGTATGCTTCAGCTTGCT 1494

Db     421 GGCAATGATGTGCCTTGACAACCGTGGGCCAGGCCTGGGCCCAGGGACTCTTCCTGTTTC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1495 GGCAATGATGTGCCTTGACAACCGTGGGCCAGGCCTGGGCCCAGGGACTCTTCCTGTTTC 1554

Db     481 AT 482
           ||
Qy    1555 AT 1556
```

```
RESULT    3
LOCUS       N56924         597 bp    mRNA            EST       22-FEB-1996
DEFINITION  yy82d03.s1 Soares_multiple_sclerosis_2NbHMSP Homo sapiens cDNA
            clone IMAGE:280037 3', mRNA sequence.
ACCESSION   N56924
NID         g1200814
VERSION     N56924.1  GI:1200814
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:692843.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: m13 -40 forward
            High quality sequence stop: 358.
FEATURES             Location/Qualifiers
     source          1. .597
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis
                     lesions from one patient was kindly provided by Dr. Kevin
                     G. Becker (NINDS/NIH). "
                     /db_xref="GDB:3898413"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:280037"
                     /clone_lib="Soares_multiple_sclerosis_2NbHMSP"
                     /sex="male"
                     /tissue_type="multiple sclerosis lesions"
                     /dev_stage="Age 46"
                     /lab_host="DH10B (ampicillin resistant)"
```

```
BASE COUNT      151 a    139 c    147 g    159 t      1 others
ORIGIN

  Query Match           19.9%;  Score 464;  DB 33;  Length 597;
  Best Local Similarity 97.6%;  Pred. No. 0.00e+00;
  Matches  490;  Conservative    0;  Mismatches   9;  Indels   3;  Gaps   3;

Db      99 TATAGAGAAATGCCTTTATAGAAAAGTAGAAACCAGTAATATTCTCTTCTCCAGCATCAC 158
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    2316 TATAGAGAAATGCCTTTATAGAAAAGTAGAAACCAGTAATATTCTCTTCTCCAGCATCAC 2257

Db     159 TAACACCAAGAGACCACCTGAGGTCTAGGTCCCCAAAGCAGATGGCTCCATAGAAAGCCC 218
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    2256 TAACACCAAGAGACCACCTGAGGTCTAGGTCCCCAAAGCAGATGGCTCCATAGAAAGCCC 2197

Db     219 CACTAACCCCTGTCTCCACATTGGGCAGTGGAAGGGTTCTGGAAAGGAAGCTCTATGGCT 278
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    2196 CACTAACCCCTGTCTCCACATTGGGCAGTGGAAGGGTTCTGGAAAGGAAGCTCTATGGCT 2137

Db     279 AGGAGCTGCCAAGGCCTCTTGAGTGTGACATCACAGGTTAGAGGCCCTGCTGAGCTGCTA 338
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    2136 AGGAGCTGCCAAGGCCTCTTGAGTGTGACATCACAGGTTAGAGGCCCTGCTGAGCTGCTA 2077

Db     339 GCACAGTGCAGTGGAGAAATGTTGCGCTCCGTCAGAAAGCTGGGCTCTTGCCATGCCAGT 398
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Cp    2076 GCACAGTGCAGTGGAGAAATGTTGCGCTCCGTCAGGAAGCTGGGCTCTTGCCATGCCAGT 2017

Db     399 AGCTTGCTTACGACCTTGGGCAAGTCACTTTTTCCTCTCTGGGGTCCTGTGTCCAAAAGG 458
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    2016 AGCTTGCTTACGACCTTGGGCAAGTCACTTTTTCCTCTCTGGGGTCCTGTGTCCAAAAGG 1957

Db     459 CTTAATGAGCCCTTTCTAATGGTTCCTTGATTGCTTTATGAGTTAGGGGCTGTGACC-TA 517
           |||||||||||||||||||||||||| |||||||||||||||||||||||||||||| ||
Cp    1956 CTTAATGAGCCCTTTCTAATGGTTCTTTGATTGCTTTATGAGTTAGGGGCTGTGACCCTA 1897

Db     518 GCCCCAGTGNCCTACAGGAAGGTGCTGACTGGGACGCCCGCAGGACCACTGCTAAATCAG 577
           ||||||||| |||||||||||||||||||  |||||||||||  | ||||||||||||
Cp    1896 GCCCCAGTGCCCTACAGGAAGGTGCTGACCTGGACGCCCGCAGACCACCTGCTAAATCAG 1837

Db     578 GCC-A-TTGGGGAGCAAGCCCT 597
           ||| | ||||||||||||||||
Cp    1836 GCCCACTTGGGGAGCAAGCCCT 1815
```

```
RESULT    4
LOCUS       AI299416       461 bp    mRNA            EST       03-DEC-1998
DEFINITION  qm89d09.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1895921 3',
            mRNA sequence.
ACCESSION   AI299416
NID         g3959070
VERSION     AI299416.1  GI:3959070
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Aug 21, 1998 this sequence version replaced.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
```



Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Mon Sep 20 20:55:46 1999; MasPar time 140.28 Seconds
1177.638 Million cell updates/sec

Tabular output not generated.

Title: >US-09-084-491A-1
Description: (1-2329) from US09084491A.seq
Perfect Score: 2329
N.A. Sequence: 1 TTACCAGAACAGCATAACAA.....CTCTATAAAAAAAAAAAAAA 2329
Comp: AATGGTCTTGTCGTATTGTT.....GAGATATTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 137044 seqs, 35465580 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-issued
1:5A_COMB 2:5B_COMB 3:5C_COMB 4:PCT9_COMB 5:backfiles1

Statistics: Mean 9.355; Variance 5.013; scale 1.866

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
c	1	64	2.7	7218	2	US-08-232-	Sequence 14, Applicati	8.57e-27
	2	43	1.8	965	3	US-08-388-	Sequence 22, Applicati	1.40e-12
c	3	41	1.8	965	3	US-08-388-	Sequence 22, Applicati	2.70e-11
c	4	40	1.7	215	1	US-08-238-	Sequence 5, Applicatio	1.17e-10
	5	34	1.5	7218	2	US-08-232-	Sequence 14, Applicati	6.19e-07
	6	32	1.4	215	1	US-08-238-	Sequence 5, Applicatio	9.74e-06
	7	28	1.2	2184	1	US-07-815-	Sequence 1, Applicatio	1.99e-03
	8	28	1.2	2288	2	US-08-290-	Sequence 4, Applicatio	1.99e-03
	9	28	1.2	2289	1	US-07-838-	Sequence 2, Applicatio	1.99e-03
	10	28	1.2	2679	5	5200340-7	Patent No. 5200340.	1.99e-03
c	11	25	1.1	66	1	US-08-471-	Sequence 144, Applicat	8.77e-02
c	12	25	1.1	69	1	US-08-471-	Sequence 142, Applicat	8.77e-02
	13	26	1.1	74	4	PCT-US95-1	Sequence 100, Applicat	2.54e-02
	14	25	1.1	74	4	PCT-US95-1	Sequence 94, Applicati	8.77e-02
c	15	25	1.1	74	4	PCT-US95-1	Sequence 94, Applicati	8.77e-02
c	16	25	1.1	75	4	PCT-US95-1	Sequence 99, Applicati	8.77e-02
	17	26	1.1	81	4	PCT-US95-1	Sequence 98, Applicati	2.54e-02
	18	25	1.1	81	4	PCT-US95-1	Sequence 92, Applicati	8.77e-02
c	19	25	1.1	81	4	PCT-US95-1	Sequence 92, Applicati	8.77e-02
c	20	25	1.1	82	4	PCT-US95-1	Sequence 97, Applicati	8.77e-02
	21	26	1.1	2296	1	US-07-750-	Sequence 18, Applicati	2.54e-02
	22	26	1.1	2497	3	US-08-643-	Sequence 12, Applicati	2.54e-02
	23	26	1.1	2753	1	US-07-854-	Sequence 1, Applicatio	2.54e-02
	24	23	1.0	65	1	US-08-471-	Sequence 145, Applicat	9.71e-01
c	25	23	1.0	65	1	US-08-471-	Sequence 145, Applicat	9.71e-01
	26	23	1.0	68	1	US-08-471-	Sequence 143, Applicat	9.71e-01
c	27	23	1.0	68	1	US-08-471-	Sequence 143, Applicat	9.71e-01
	28	23	1.0	69	1	US-08-471-	Sequence 142, Applicat	9.71e-01
c	29	24	1.0	74	4	PCT-US95-1	Sequence 100, Applicat	2.96e-01
c	30	24	1.0	81	4	PCT-US95-1	Sequence 98, Applicati	2.96e-01
	31	24	1.0	82	4	PCT-US95-1	Sequence 97, Applicati	2.96e-01
	32	24	1.0	242	2	US-08-273-	Sequence 1, Applicatio	2.96e-01
	33	23	1.0	2033	2	US-08-448-	Sequence 14, Applicati	9.71e-01
	34	23	1.0	2033	1	US-08-148-	Sequence 14, Applicati	9.71e-01
	35	23	1.0	2219	1	US-07-882-	Sequence 1, Applicatio	9.71e-01
	36	23	1.0	2219	1	US-08-184-	Sequence 2, Applicatio	9.71e-01
	37	23	1.0	2219	1	US-08-184-	Sequence 1, Applicatio	9.71e-01
	38	23	1.0	2232	2	US-08-334-	Sequence 1, Applicatio	9.71e-01
	39	23	1.0	2232	4	PCT-US95-1	Sequence 1, Applicatio	9.71e-01
	40	23	1.0	2262	1	US-07-882-	Sequence 7, Applicatio	9.71e-01
c	41	24	1.0	2380	3	US-08-572-	Sequence 3, Applicatio	2.96e-01
	42	24	1.0	3603	1	US-08-646-	Sequence 15, Applicati	2.96e-01
	43	24	1.0	3603	1	US-08-188-	Sequence 15, Applicati	2.96e-01
c	44	23	1.0	4673	1	US-07-638-	Sequence 1, Applicatio	9.71e-01
	45	23	1.0	6100	1	US-08-184-	Sequence 6, Applicatio	9.71e-01

ALIGNMENTS

```
RESULT    1
ID   US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.
AC   xxxxxx
DT
DE   Sequence 14, Application US/08232463
CC   Sequence 14, Application US/08232463
CC   Patent No. 5670367
CC    GENERAL INFORMATION:
CC      APPLICANT:  DORNER, F.
CC      APPLICANT:  SCHEIFLINGER, F.
CC      APPLICANT:  FALKNER, F. G.
CC      TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
CC      NUMBER OF SEQUENCES:  52
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Foley & Lardner
CC        STREET:  1800 Diagonal Road, Suite 500
CC        CITY:  Alexandria
CC        STATE:  VA
CC        COUNTRY:  USA
CC        ZIP:  22313-0299
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/232,463
CC        FILING DATE:
CC        CLASSIFICATION:  435
CC      PRIOR APPLICATION DATA:
CC        APPLICATION NUMBER:  US/07/935,313
CC        FILING DATE:
CC        APPLICATION NUMBER:  EP 91 114 300.6
CC        FILING DATE:  26-AUG-1991
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  BENT, Stephen A.
CC        REGISTRATION NUMBER:  29,768
CC        REFERENCE/DOCKET NUMBER:  30472/114 IMMU
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (703)836-9300
CC        TELEFAX:  (703)683-4109
CC        TELEX:  899149
CC    INFORMATION FOR SEQ ID NO:  14:
CC      SEQUENCE CHARACTERISTICS:
```

```
CC        LENGTH:  7218 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      IMMEDIATE SOURCE:
CC        CLONE:  pTZgpt-F1s
SQ   SEQUENCE 7218 BP; 1944 A; 1491 C; 1486 G; 1929 T; 368 OTHER.

  Query Match            2.7%;  Score 64;  DB 2;  Length 7218;
  Best Local Similarity   1.0%;  Pred. No. 8.57e-27;
  Matches     3;  Conservative  184;  Mismatches 123;  Indels   0;  Gaps   0;

Db   1064 GATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1123
          |||:: :: : :::::::::: : : :::::: :: : :: :::::::   ::::::::
Cp    771 GATTCGCTGCATCTCCCTCTCACATACTTTCTGATCATGCTGTTCTTTCAAATCCTTCCC 712

Db   1124 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1183
          :::::: :     :  :::    :  : : ::  ::::  :  :  :  :  :: :: :: :
Cp    711 CCTCTTGTAGGAGTAGCCCAAGATGATGCCAGCTCCGATGGCAATGATGATCACCATCAT 652

Db   1184 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1243
            :  ::: :  : :  :::    :::::   :::::::::::::    ::: :::
Cp    651 GGTAATGCCCAGCACGTAGCCCAGAGTTCCCAGGTCCTTTTTCTCCTTGGAGTTCATCCG 592

Db   1244 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1303
          : ::: ::  ::  :::: :: ::  :: : :  :: :: :::: ::::   :   :
Cp    591 CACCCGCTGGCTGATCCCAATCACTGGCTGCACAGCTGCCGCCTCACTCCGAGCGGGCAG 532

Db   1304 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1363
            : ::  :     :   : ::: : :::: ::: : :::  :::::   : ::::::
Cp    531 GGCGTTGGCAGGAGCGAACACCTGCACCTCATCTGCACCTGGCCCTTCAGACGCTTCCTG 472

Db   1364 YYYYYYYYYY 1373
            ::::: ::
Cp    471 GATTTCTGTC 462


RESULT    2
ID   US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.
AC   xxxxxx
DT
DE   Sequence 22, Application US/08388672A
CC   Sequence 22, Application US/08388672A
CC   Patent No. 5795961
CC    GENERAL INFORMATION:
CC      APPLICANT:  Wallace, T. Paul
CC      APPLICANT:  Harris, William J.
CC      APPLICANT:  Carr, Frank J.
CC      APPLICANT:  Old, Lloyd J.
CC      APPLICANT:  Welt, Sydney
CC      APPLICANT:  Kitamura, Kunio
CC      TITLE OF INVENTION:  Recombinant Human Anti-Lewis B
CC      TITLE OF INVENTION:  Antibodies
CC      NUMBER OF SEQUENCES:  25
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Felfe and Lynch
CC        STREET:  805 Third Avenue
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10022
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/388,672A
CC        FILING DATE:  14-FEB-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Hanson, No. 5795961man D.
CC        REGISTRATION NUMBER:  30,946
CC        REFERENCE/DOCKET NUMBER:  LUD 5409
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212-688-9200
CC        TELEFAX:  212-838-3884
CC    INFORMATION FOR SEQ ID NO:  22:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  965 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  unknown
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.

  Query Match            1.8%;  Score 43;  DB 3;  Length 965;
  Best Local Similarity  18.1%;  Pred. No. 1.40e-12;
  Matches    33;  Conservative   84;  Mismatches  62;  Indels   3;  Gaps   3;

Db    774 TGGGTKKGURHUVHVSGGVRSTSTCTASDYTTSYWGWVRGRGWGDYGGGYTNYNGKRGRV 833
          |||| ::| :::::     :  :||| :::  :: |::: :   ::|  :|   |:  ::
Qy   1381 TGGGATGGAACTCCCTACTGACCTCT-GACAACTGGAAACGAGTTTGTACTGAAGTCAGA 1439

Db    834 TMADTSSNSRSSVTAADTAVYYCVRGRSYDSDGGDYWGGTTVTVSSHUVKDMTSSSSASV 893
           : :|::   ::: | :  ::: ::|   :::  :::||| :|:: : ::  |:: : ::
Qy   1440 ACTTTGGGTTGGGAATGAGATCTAGGTTGTGGCTGCTGGT-ATGCTTCAGCTTGCTGGCA 1498

Db    894 GDRVTTCRSSTTHGNGNTYYWYKGKAKYRVSNRSGVSRSGSGSGTDYTTSSDATYYCGTH 953
           :::|  :::|| |    :: ::|:   :::  :|:: : :|:  ::||::: |::| |:
Qy   1499 ATGATGTGCCTT-GACAACCGTGGGCCAGGCCTGGGCCCAGGGACTCTTCCTGTTTCATA 1557

Db    954 AR 955
          |:
Qy   1558 AG 1559


RESULT    3
ID   US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.
AC   xxxxxx
DT
DE   Sequence 22, Application US/08388672A
CC   Sequence 22, Application US/08388672A
CC   Patent No. 5795961
CC    GENERAL INFORMATION:
CC      APPLICANT:  Wallace, T. Paul
CC      APPLICANT:  Harris, William J.
CC      APPLICANT:  Carr, Frank J.
CC      APPLICANT:  Old, Lloyd J.
CC      APPLICANT:  Welt, Sydney
CC      APPLICANT:  Kitamura, Kunio
CC      TITLE OF INVENTION:  Recombinant Human Anti-Lewis B
CC      TITLE OF INVENTION:  Antibodies
CC      NUMBER OF SEQUENCES:  25
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Felfe and Lynch
CC        STREET:  805 Third Avenue
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10022
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/388,672A
CC        FILING DATE:  14-FEB-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Hanson, No. 5795961man D.
CC        REGISTRATION NUMBER:  30,946
CC        REFERENCE/DOCKET NUMBER:  LUD 5409
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212-688-9200
```

```
CC        TELEFAX:  212-838-3884
CC    INFORMATION FOR SEQ ID NO:  22:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  965 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  unknown
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.

  Query Match            1.8%;  Score 41;  DB 3;  Length 965;
  Best Local Similarity  14.4%;  Pred. No. 2.70e-11;
  Matches    27;  Conservative    84;  Mismatches  76;  Indels   0;  Gaps   0;

Db    776 GGTKKGURHUVHVSGGVRSTSTCTASDYTTSYWGWVRGRGWGDYGGGYTNYNGKRGRVTM 835
          || ::|::: ::::  ::: :||| ::: |::: ::    : : | |   : |: |::
Cp    847 GGCTGGTGTGGACCACGACAGTCTTCTCATCCACAATCTCACAGGTGGGGTTGGTGAAGG 788

Db    836 ADTSSNSRSSVTAADTAVYYCVRGRSYDSDGGDYWGGTTVTVSSHUVKDMTSSSSASVGD 895
           :  :  :::: |:| :::|:  :: :::   :     :|:: ::::::|:  :  : :
Cp    787 CAGACAAGGGCAGAGTGATTCGCTGCATCTCCCTCTCACATACTTTCTGATCATGCTGTT 728

Db    896 RVTTCRSSTTHGNGNTYYWYKGKAKYRVSNRSGVSRSGSGSGTDYTTSSDATYYCGTHAR 955
            |||:  | :     ::::: :   ::: :: ::: |  : |::  :::     | :|:
Cp    727 CTTTCAAATCCTTCCCCCTCTTGTAGGAGTAGCCCAAGATGATGCCAGCTCCGATGGCAA 668

Db    956 TGGTKVK 962
          || |:::
Cp    667 TGATGAT 661


RESULT    4
ID   US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.
AC   xxxxxx
DT
DE   Sequence 5, Application US/08238163
CC   Sequence 5, Application US/08238163
CC   Patent No. 5569830
CC    GENERAL INFORMATION:
CC      APPLICANT:  BENNETT, Alan
CC      APPLICANT:  LABAVITCH, John M.
CC      APPLICANT:  POWELL, Ann
CC      APPLICANT:  STOTZ, Henrik
CC      TITLE OF INVENTION:  PLANT INHIBITORS OF FUNGAL
CC      TITLE OF INVENTION:  POLYGALACTURONASES AND THEIR USE TO CONTROL FUNGAL DISEASE
CC      NUMBER OF SEQUENCES:  24
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Townsend and Townsend Khourie and Crew
CC        STREET:  Steuart Street Tower, One Market Plaza
CC        CITY:  San Francisco
CC        STATE:  California
CC        COUNTRY:  US
CC        ZIP:  94105-1493
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/238,163
CC        FILING DATE:  03-MAY-1994
CC        CLASSIFICATION:  800
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Bastian, Kevin L.
CC        REGISTRATION NUMBER:  34,774
CC        REFERENCE/DOCKET NUMBER:  2307E-540
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (415) 543-9600
CC        TELEFAX:  (415) 543-5043
CC    INFORMATION FOR SEQ ID NO:  5:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  215 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  protein
CC      FEATURE:
CC        NAME/KEY:  misc_feature
CC        LOCATION:  1..215
CC        OTHER INFORMATION:  /standard_name= "Deduced amino acid
CC        OTHER INFORMATION:  sequence of PGIP from bean."
SQ   SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

  Query Match            1.7%;  Score 40;  DB 1;  Length 215;
  Best Local Similarity  14.9%;  Pred. No. 1.17e-10;
  Matches    26;  Conservative    72;  Mismatches  76;  Indels   1;  Gaps   1;

Db      9 SSVVSRTASCNDKAKKDGNTTSSWT-TDCCNRTWGVCDTDTTYRVNNDSGHNKYSSANYN 67
          :::: : |:  :: :::  ||:::  |: | :|:|: : :|  ::  :: :   ::| :
Cp   1005 CCAGAACACAAAGTGGTAGTTCCTCCTAGCTGTAGGAGGGTGGGCTGTCCTGCACCAGTG 946

Db     68 YGGNNVGAAKTHYYTHTNVSGADSKTVTDSYNASGTSSSNGGTDGNRSGADSYGSSKTAM 127
           :    :   : :::|   :   ::: : ::: |:| ::: || :   :|  ::|::  |
Cp    945 TCTGCATGGGCTCCTGCCCACTGGGGGGGCTCAGGCCCCAGGAGTCCCGGCCTGGCCCAT 886

Db    128 TSRNRTGKTANNAVDSRNMGDASVGSDKNTKKHAKNSADGKVGSKNNGDRNNRYG 182
            : : |:|    : :  : :|::|     :: |: : :|  |::  |::  ::|
Cp    885 AAGGGGGGTGCTGCCCTCCTGAGGGTCAACTGGAGTCTGGCTGGTGTGGACCACG 831


RESULT    5
ID   US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.
AC   xxxxxx
DT
DE   Sequence 14, Application US/08232463
CC   Sequence 14, Application US/08232463
CC   Patent No. 5670367
CC    GENERAL INFORMATION:
CC      APPLICANT:  DORNER, F.
CC      APPLICANT:  SCHEIFLINGER, F.
CC      APPLICANT:  FALKNER, F. G.
CC      TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
CC      NUMBER OF SEQUENCES:  52
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Foley & Lardner
CC        STREET:  1800 Diagonal Road, Suite 500
CC        CITY:  Alexandria
CC        STATE:  VA
CC        COUNTRY:  USA
CC        ZIP:  22313-0299
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/232,463
CC        FILING DATE:
CC        CLASSIFICATION:  435
CC      PRIOR APPLICATION DATA:
CC        APPLICATION NUMBER:  US/07/935,313
CC        FILING DATE:
CC        APPLICATION NUMBER:  EP 91 114 300.6
CC        FILING DATE:  26-AUG-1991
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  BENT, Stephen A.
CC        REGISTRATION NUMBER:  29,768
CC        REFERENCE/DOCKET NUMBER:  30472/114 IMMU
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (703)836-9300
CC        TELEFAX:  (703)683-4109
CC        TELEX:  899149
CC    INFORMATION FOR SEQ ID NO:  14:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  7218 base pairs
CC        TYPE:  nucleic acid
```

```
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      IMMEDIATE SOURCE:
CC        CLONE:  pTZgpt-Fls
SQ   SEQUENCE 7218 BP; 1944 A; 1491 C; 1486 G; 1929 T; 368 OTHER.

  Query Match           1.5%;  Score 34;  DB 2;  Length 7218;
  Best Local Similarity   1.5%;  Pred. No. 6.19e-07;
  Matches     4;  Conservative  148;  Mismatches 118;  Indels   0;  Gaps   0;

Db   1062 GCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1121
          |||| :: :::: ::::: ::: ::::: ::  :::: ::: :    :: :   :
Qy    765 GCGAATCACTCTGCCCTTGTCTGCCTTCACCAACCCCACCTGTGAGATTGTGGATGAGAA 824

Db   1122 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1181
           :: :: :  ::: : ::  ::   :::: ::  ::::       :  : :::::::
Qy    825 GACTGTCGTGGTCCACACCAGCCAGACTCCAGTTGACCCTCAGGAGGGCAGCACCCCCCT 884

Db   1182 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1241
           : :   ::   ::    :::::   :::   :::::::  :   :     ::: : :
Qy    885 TATGGGCCAGGCCGGGACTCCTGGGGCCTGAGCCCCCCCAGTGGGCAGGAGCCCATGCAG 944

Db   1242 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1301
            : ::  : :    :  ::: :::::::: :  ::        :: :: :::: : ::::
Qy    945 ACACTGGTGCAGGACAGCCCACCCTCCTACAGCTAGGAGGAACTACCACTTTGTGTTCTG 1004

Db   1302 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1331
            ::    :::: :: ::::::: :::::::
Qy   1005 GTTAAAACCCTACCACTCCCCCGCTTTTTT 1034


RESULT    6
ID   US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.
AC   xxxxxx
DT
DE   Sequence 5, Application US/08238163
CC   Sequence 5, Application US/08238163
CC   Patent No. 5569830
CC    GENERAL INFORMATION:
CC      APPLICANT:  BENNETT, Alan
CC      APPLICANT:  LABAVITCH, John M.
CC      APPLICANT:  POWELL, Ann
CC      APPLICANT:  STOTZ, Henrik
CC      TITLE OF INVENTION:  PLANT INHIBITORS OF FUNGAL
CC      TITLE OF INVENTION:  POLYGALACTURONASES AND THEIR USE TO CONTROL FUNGAL DISEASE
CC      NUMBER OF SEQUENCES:  24
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Townsend and Townsend Khourie and Crew
CC        STREET:  Steuart Street Tower, One Market Plaza
CC        CITY:  San Francisco
CC        STATE:  California
CC        COUNTRY:  US
CC        ZIP:  94105-1493
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/238,163
CC        FILING DATE:  03-MAY-1994
CC        CLASSIFICATION:  800
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Bastian, Kevin L.
CC        REGISTRATION NUMBER:  34,774
CC        REFERENCE/DOCKET NUMBER:  2307E-540
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (415) 543-9600
CC        TELEFAX:  (415) 543-5043
CC     INFORMATION FOR SEQ ID NO:  5:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  215 base pairs
CC         TYPE:  nucleic acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
CC       FEATURE:
CC         NAME/KEY:  misc_feature
CC         LOCATION:  1..215
CC         OTHER INFORMATION:  /standard_name= "Deduced amino acid
CC         OTHER INFORMATION:  sequence of PGIP from bean."
SQ   SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

  Query Match           1.4%;  Score 32;  DB 1;  Length 215;
  Best Local Similarity  15.5%;  Pred. No. 9.74e-06;
  Matches    22;  Conservative   57;  Mismatches  62;  Indels   1;  Gaps   1;

Db     41 WGVCDTDTTYRVNNDSGHNKYSSANYNYGGNNVGAAKTHYYTHTN-VSGADSKTVTDSYN 99
          :|:  |:  :::  :: : : ::|   :||  :  |: ::: :   ::||::: : : :
Qy    629 TGGGCTACGTGCTGGGCATTACCATGATGGTGATCATCATTGCCATCGGAGCTGGCATCA 688

Db    100 ASGTSSSNGGTDGNRSGADSYGSSKTAMTSRNRTGKTANNAVDSRNMGDASVGSDKNTKK 159
           : |:::    :     |:: |::: |: ::  ||  |  |:::   |: ::| :  | :
Qy    689 TCTTGGGCTACTCCTACAAGAGGGGGAAGGATTTGAAAGAACAGCATGATCAGAAAGTAT 748

Db    160 HAKNSADGKVGSKNNGDRNNRY 181
            : :|:|::| :   :   ::
Qy    749 GTGAGAGGGAGATGCAGCGAAT 770


RESULT    7
ID   US-07-815-333A-1 STANDARD; DNA; UNC; 2184 BP.
AC   xxxxxx
DT
DE   Sequence 1, Application US/07815333A
CC   Sequence 1, Application US/07815333A
CC   Patent No. 5342831
CC    GENERAL INFORMATION:
CC      APPLICANT:  Nakamura, Toshikazu
CC      APPLICANT:  Matsumoto, Kunio
CC      TITLE OF INVENTION:  EPITHELIOCYTE GROWTH ACCELERATOR
CC      NUMBER OF SEQUENCES:  2
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Leydig, Voit & Mayer
CC        STREET:  Two Prudential Plaza, Suite 4900
CC        CITY:  Chicago
CC        STATE:  Illinois
CC        COUNTRY:  United States of America
CC        ZIP:  60601
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/07/815,333A
CC        FILING DATE:  19911227
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Green, Robert F.
CC        REGISTRATION NUMBER:  27555
CC        REFERENCE/DOCKET NUMBER:  44069
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (312) 616-5600
CC        TELEFAX:  (312) 616-5700
CC        TELEX:  25-3533
CC     INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  2184 base pairs
CC         TYPE:  NUCLEIC ACID
CC         STRANDEDNESS:  double
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  cDNA
CC       FEATURE:
CC         NAME/KEY:  CDS
CC         LOCATION:  1..2184
```

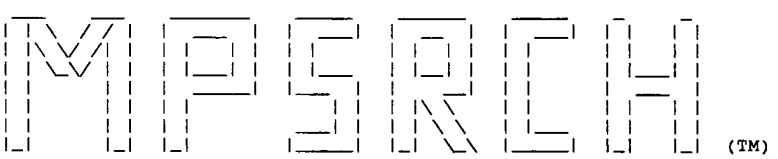
```
SQ   SEQUENCE 2184 BP; 673 A; 455 C; 500 G; 556 T; 0 OTHER.

  Query Match             1.2%;  Score 28;  DB 1;  Length 2184;
  Best Local Similarity  75.9%;  Pred. No. 1.99e-03;
  Matches    41;  Conservative    0;  Mismatches  13;  Indels   0;  Gaps   0;

Db   1306 AATGAGAATTACTGCCGAAATCCAGATGATGATGCTCATGGACCCTGGTGCTAC 1359
          ||| | | |||||||||||||| || || || ||  | |  || ||||||||||||
Qy    319 AATCACAGTTACTGCCGAAACCCGGACGAGGACCCGCGCGGGCCCTGGTGCTAC 372


RESULT    8
ID   US-08-290-937B-4 STANDARD; DNA; UNC; 2288 BP.
AC   xxxxxx
DT
DE   Sequence 4, Application US/08290937B
CC   Sequence 4, Application US/08290937B
CC   Patent No. 5648233
CC    GENERAL INFORMATION:
CC      APPLICANT:  YAMAGUCHI, KYOJI
CC      APPLICANT:  SHIMA, NOBUYUKI
CC      APPLICANT:  MURAKAMI, AKIHIKO
CC      APPLICANT:  GOTO, MASAAKI
CC      APPLICANT:  TSUDA, EISUKE
CC      APPLICANT:  MASUNAGA, HIROAKI
CC      APPLICANT:  TAKAHIRA, REIKO
CC      APPLICANT:  OOGAKI, FUMIKO
CC      APPLICANT:  UEDA, MASATSUGU
CC      APPLICANT:  HIGASHIO, KANJI
CC      TITLE OF INVENTION:  MODIFIED TCF
CC      NUMBER OF SEQUENCES:  13
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Testa, Hurwitz & Thibeault
CC        STREET:  125 High St.
CC        CITY:  Boston
CC        STATE:  MA
CC        COUNTRY:  USA
CC        ZIP:  02110
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/290,937B
CC        FILING DATE:  19-AUG-1994
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  CAMPBELL, PAULA A.
CC        REGISTRATION NUMBER:  32,503
CC        REFERENCE/DOCKET NUMBER:  FJN-022
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (617) 248-7000
CC        TELEFAX:  (617) 248-7100
CC     INFORMATION FOR SEQ ID NO:  4:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  2288 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  cDNA
SQ   SEQUENCE 2288 BP; 702 A; 491 C; 524 G; 571 T; 0 OTHER.

  Query Match             1.2%;  Score 28;  DB 2;  Length 2288;
  Best Local Similarity  75.9%;  Pred. No. 1.99e-03;
  Matches    41;  Conservative    0;  Mismatches  13;  Indels   0;  Gaps   0;

Db   1367 AATGAGAATTACTGCCGAAATCCAGATGATGATGCTCATGGACCCTGGTGCTAC 1420
          ||| | | |||||||||||||| || || || ||  | |  || ||||||||||||
Qy    319 AATCACAGTTACTGCCGAAACCCGGACGAGGACCCGCGCGGGCCCTGGTGCTAC 372


RESULT    9
ID   US-07-838-410-2 STANDARD; DNA; UNC; 2289 BP.
AC   xxxxxx
DT
DE   Sequence 2, Application US/07838410
CC   Sequence 2, Application US/07838410
CC   Patent No. 5328836
CC    GENERAL INFORMATION:
CC      APPLICANT:  SHIMA, NOBUYUKI
CC      APPLICANT:  HIGASHIO, KANJI
CC      APPLICANT:  NAGAO, MASAYA
CC      APPLICANT:  OOGAKI, FUMIKO
CC      APPLICANT:  TAKAOKA, HIROAKI
CC      APPLICANT:  TSUDA, EISUKE
CC      TITLE OF INVENTION:  PLASMIDS CONTAINING DNA ENCODING THE
CC      TITLE OF INVENTION:  AMINO ACID SEQUENCE OF TCF-II, TRANSFORMED CELLS WITH TH
CC      TITLE OF INVENTION:  PLASMIDS, AND PRODUCTION MEDTHOD OF BIOLOGICALLY ACTIVE
CC      TITLE OF INVENTION:  SUBSTANCE USING THE TRANSFORMED CELLS
CC      NUMBER OF SEQUENCES:  11
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  MILLEN, WHITE, ZELANO & BRANIGAN, P.C.
CC        STREET:  2200 CLARENDON BLVD.
CC        CITY:  ARLINGTON
CC        STATE:  VA
CC        COUNTRY:  USA
CC        ZIP:  22201
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/07/838,410
CC        FILING DATE:  19920311
CC        CLASSIFICATION:  435
CC      PRIOR APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/JP91/00942
CC        FILING DATE:  15-JUL-1991
CC      PRIOR APPLICATION DATA:
CC        APPLICATION NUMBER:  JP 185852-1990
CC        FILING DATE:  13-JUL-1990
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  WHITE, JOHN L.
CC        REGISTRATION NUMBER:  17,746
CC        REFERENCE/DOCKET NUMBER:  WAK 110
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  703-243-6333
CC        TELEFAX:  703-243-6410
CC        TELEX:  64191
CC     INFORMATION FOR SEQ ID NO:  2:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  2289 base pairs
CC        TYPE:  NUCLEIC ACID
CC        STRANDEDNESS:  double
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
CC      ORIGINAL SOURCE:
CC        ORGANISM:  Homo sapiens
CC        CELL TYPE:  Fibroblast
SQ   SEQUENCE 2289 BP; 703 A; 491 C; 523 G; 572 T; 0 OTHER.

  Query Match             1.2%;  Score 28;  DB 1;  Length 2289;
  Best Local Similarity  75.9%;  Pred. No. 1.99e-03;
  Matches    41;  Conservative    0;  Mismatches  13;  Indels   0;  Gaps   0;

Db   1368 AATGAGAATTACTGCCGAAATCCAGATGATGATGCTCATGGACCCTGGTGCTAC 1421
          ||| | | |||||||||||||| || || || ||  | |  || ||||||||||||
Qy    319 AATCACAGTTACTGCCGAAACCCGGACGAGGACCCGCGCGGGCCCTGGTGCTAC 372


RESULT   10
ID   5200340-7 STANDARD; DNA; UNC; 2902 BP.
AC   xxxxxx
DT   01-JAN-1900
```

```
DE   Patent No. 5200340.
CC   Patent No. 5200340
CC       APPLICANT: FOSTER, DONALD C.;MULVIHILL, EILEEN R.;O'HARA,
CC   PATRICK J.;PINGEL, KURT;YOSHITAKE, SHINJI
CC       TITLE OF INVENTION: THROMBIN-ACTIVATED TISSUE PLASMINOGEN
CC   ACTIVATORS
CC       NUMBER OF SEQUENCES: 34
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/53,412
CC         FILING DATE: 22-MAY-1987
CC   SEQ ID NO:7:
CC         LENGTH: 2679
SQ   Sequence 2902 BP; 750 A; 659 C; 675 G; 595 T; 223 other;

  Query Match             1.2%;  Score 28;  DB 5;  Length 2679;
  Best Local Similarity  81.8%;  Pred. No. 1.99e-03;
  Matches   36;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Db     472 TACTGCAGAAATCCAGACAACGATCCGCAGGGGCCCTGGTGCTA 515
           |||||| |||| || ||| | || ||||   |||||||||||||
Qy     328 TACTGCCGAAACCCGGACGAGGACCCGCGCGGGCCCTGGTGCTA 371


RESULT   11
ID   US-08-471-052A-144 STANDARD; DNA; UNC; 66 BP.
AC   xxxxxx
DT
DE   Sequence 144, Application US/08471052A
CC   Sequence 144, Application US/08471052A
CC   Patent No. 5625033
CC    GENERAL INFORMATION:
CC      APPLICANT:  Kay, B. K.
CC      APPLICANT:  Fowlkes, D. M.
CC      TITLE OF INVENTION:  Totally Synthetic Affinity Reagents
CC      NUMBER OF SEQUENCES:  166
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10036-2711
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/471,052A
CC        FILING DATE:  06-JUNE-1995
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-179
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212 790-9090
CC        TELEFAX:  212 869-8864/9741
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  144:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  66 bases
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA
SQ   SEQUENCE 66 BP; 2 A; 3 C; 4 G; 2 T; 55 OTHER.

  Query Match             1.1%;  Score 25;  DB 1;  Length 66;
  Best Local Similarity  12.5%;  Pred. No. 8.77e-02;
  Matches    8;  Conservative   18;  Mismatches  38;  Indels    0;  Gaps    0;

Db       3 GAGNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBCC 62
           |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  |
Cp     865 GAGGGTCAACTGGAGTCTGGCTGGTGTGGACCACGACAGTCTTCTCATCCACAATCTCAC 806

Db      63 AGGT 66
           ||||
Cp     805 AGGT 802


RESULT   12
ID   US-08-471-052A-142 STANDARD; DNA; UNC; 69 BP.
AC   xxxxxx
DT
DE   Sequence 142, Application US/08471052A
CC   Sequence 142, Application US/08471052A
CC   Patent No. 5625033
CC    GENERAL INFORMATION:
CC      APPLICANT:  Kay, B. K.
CC      APPLICANT:  Fowlkes, D. M.
CC      TITLE OF INVENTION:  Totally Synthetic Affinity Reagents
CC      NUMBER OF SEQUENCES:  166
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10036-2711
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/471,052A
CC        FILING DATE:  06-JUNE-1995
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-179
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212 790-9090
CC        TELEFAX:  212 869-8864/9741
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  142:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  69 bases
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA
SQ   SEQUENCE 69 BP; 2 A; 4 C; 6 G; 2 T; 55 OTHER.

  Query Match             1.1%;  Score 25;  DB 1;  Length 69;
  Best Local Similarity  12.5%;  Pred. No. 8.77e-02;
  Matches    8;  Conservative   18;  Mismatches  38;  Indels    0;  Gaps    0;

Db       6 GAGNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBCC 65
           |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  |
Cp     865 GAGGGTCAACTGGAGTCTGGCTGGTGTGGACCACGACAGTCTTCTCATCCACAATCTCAC 806

Db      66 AGGT 69
           ||||
Cp     805 AGGT 802


RESULT   13
ID   PCT-US95-11934-100 STANDARD; DNA; UNC; 74 BP.
AC   xxxxxx
DT
DE   Sequence 100, Application PC/TUS9511934
CC   Sequence 100, Application PC/TUS9511934
CC    GENERAL INFORMATION:
```

```
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  100:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  74 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 74 BP; 6 A; 6 C; 1 G; 1 T; 60 OTHER.

  Query Match             1.1%;  Score 26;  DB 4;  Length 74;
  Best Local Similarity  11.4%;  Pred. No. 2.54e-02;
  Matches     8;  Conservative   20;  Mismatches  42;  Indels   0;  Gaps   0;

Db       3 AGAVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNN 62
           |||:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy     108 AGACGAAGAGGCAAGGATGCTGTTGGCCTGGGTACAAGCATTCCTCGTCAGCAACATGCT 167

Db      63 VNNACCACAA 72
           :  | || ||
Qy     168 CCTAGCAGAA 177


RESULT   14
ID   PCT-US95-11934-94 STANDARD; DNA; UNC; 74 BP.
AC   xxxxxx
DT
DE   Sequence 94, Application PC/TUS9511934
CC   Sequence 94, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  94:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  74 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 74 BP; 3 A; 4 C; 3 G; 1 T; 63 OTHER.

  Query Match             1.1%;  Score 25;  DB 4;  Length 74;
  Best Local Similarity   7.5%;  Pred. No. 8.77e-02;
  Matches     5;  Conservative   20;  Mismatches  42;  Indels   0;  Gaps   0;

Db       5 GNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBN 64
           |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy     212 GCCACCTGTACCGGGAGGACCAGACCTCCCCCGCGCCGGGCCTCCGCTGCCTCAACTGGC 271

Db      65 NBNACGC 71
           :.||||
Qy     272 TGGACGC 278


RESULT   15
ID   PCT-US95-11934-94 STANDARD; DNA; UNC; 74 BP.
AC   xxxxxx
DT
DE   Sequence 94, Application PC/TUS9511934
CC   Sequence 94, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  94:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  74 base pairs
CC        TYPE:  nucleic acid
```

```
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 74 BP; 3 A; 4 C; 3 G; 1 T; 63 OTHER.

  Query Match             1.1%;  Score 25;  DB 4;  Length 74;
  Best Local Similarity   8.7%;  Pred. No. 8.77e-02;
  Matches     6;  Conservative   20;  Mismatches  43;  Indels   0;  Gaps   0;

Db       5 GNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBN 64
           |   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Cp     896 GCCTGGCCCATAAGGGGGGTGCTGCCCTCCTGAGGGTCAACTGGAGTCTGGCTGGTGTGG 837

Db      65 NBNACGCCA 73
            : ||| ||
Cp     836 ACCACGACA 828


Search completed: Mon Sep 20 21:00:25 1999
Job time : 279 secs.
```

****



****

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Mon Sep 20 20:38:15 1999; MasPar time 499.24 Seconds
999.603 Million cell updates/sec

Tabular output not generated.

Title: >US-09-084-491A-1
Description: (1-2329) from US09084491A.seq
Perfect Score: 2329
N.A. Sequence: 1 TTACCAGAACAGCATAACAA.....CTCTATAAAAAAAAAAAAAA 2329
Comp: AATGGTCTTGTCGTATTGTT.....GAGATATTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60

Statistics: Mean 9.831; Variance 6.007; scale 1.637

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	2329	100.0	2329	60	V99636	Human tissue plasmino	0.00e+00
	2	270	11.6	399	60	V99640	Human t-PALP-related	3.38e-159
	3	267	11.5	472	60	V99642	Human t-PALP-related	3.88e-157
	4	238	10.2	250	60	V99637	Human t-PALP-related	2.82e-137
c	5	235	10.1	247	60	V99638	Human t-PALP-related	3.16e-135
	6	191	8.2	461	60	V99639	Human t-PALP-related	2.63e-105
	7	150	6.4	334	60	V99641	Human t-PALP-related	9.67e-78
	8	74	3.2	291	60	V99643	Human t-PALP-related	2.54e-28
	9	44	1.9	204	1	N81164	Base substituted E.co	2.32e-10
	10	42	1.8	91	9	Q51746	Oligonucleotide probe	3.07e-09
c	11	43	1.8	91	9	Q51746	Oligonucleotide probe	8.48e-10
c	12	40	1.7	204	1	N81164	Base substituted E.co	3.94e-08
c	13	38	1.6	114	12	Q70468	Generic DNA sequence	4.87e-07
c	14	37	1.6	114	12	Q70469	Generic DNA sequence	1.69e-06
	15	36	1.5	114	12	Q70465	Generic DNA sequence	5.78e-06
	16	36	1.5	114	12	Q70468	Generic DNA sequence	5.78e-06
	17	36	1.5	114	12	Q70467	Generic DNA sequence	5.78e-06
	18	36	1.5	114	12	Q70469	Generic DNA sequence	5.78e-06
	19	35	1.5	114	12	Q70470	Generic DNA sequence	1.96e-05
	20	34	1.5	114	12	Q70466	Generic DNA sequence	6.56e-05
c	21	35	1.5	114	12	Q70467	Generic DNA sequence	1.96e-05
c	22	34	1.5	114	12	Q70466	Generic DNA sequence	6.56e-05
c	23	34	1.5	114	12	Q70470	Generic DNA sequence	6.56e-05
c	24	34	1.5	114	12	Q70473	Generic DNA sequence	6.56e-05
c	25	34	1.5	114	12	Q70465	Generic DNA sequence	6.56e-05
c	26	36	1.5	140	32	T76368	Human IL-8 receptor-a	5.78e-06
c	27	36	1.5	190	32	T76452	Chymase antisense oli	5.78e-06
	28	32	1.4	114	12	Q70473	Generic DNA sequence	7.10e-04
	29	32	1.4	114	12	Q70472	Generic DNA sequence	7.10e-04
c	30	32	1.4	114	12	Q70472	Generic DNA sequence	7.10e-04
c	31	32	1.4	114	12	Q70471	Generic DNA sequence	7.10e-04
	32	32	1.4	130	47	V48104	Randomised Pool oligo	7.10e-04
	33	32	1.4	178	32	T76405	Human endothelin-1 an	7.10e-04
c	34	32	1.4	178	32	T76405	Human endothelin-1 an	7.10e-04
	35	31	1.3	114	12	Q70471	Generic DNA sequence	2.29e-03
c	36	30	1.3	130	47	V48104	Randomised Pool oligo	7.26e-03
	37	30	1.3	160	47	V48098	Oligonucleotide LP160	7.26e-03
c	38	31	1.3	160	47	V48098	Oligonucleotide LP160	2.29e-03
	39	31	1.3	190	32	T76452	Chymase antisense oli	2.29e-03
	40	31	1.3	1140	1	Q02301	cPA-P2 Hybrid plasmin	2.29e-03
c	41	29	1.2	172	32	T76363	Human interleukin 8 a	2.27e-02
c	42	29	1.2	984	17	Q94336	Degenerate Alteromona	2.27e-02
	43	28	1.2	2187	8	Q47833	Competative inhibitor	6.97e-02
	44	28	1.2	2187	8	Q47832	Competative inhibitor	6.97e-02
	45	28	1.2	2679	35	T89686	Plasminogen encoding	6.97e-02

ALIGNMENTS

RESULT 1
ID V99636 standard; DNA; 2329 BP.
AC V99636;
DT 29-MAR-1999 (first entry)
DE Human tissue plasminogen activator-like protease t-PALP DNA.
KW Tissue plasminogen activator-like protease; t-PALP; human;
KW circulatory system-related disorder; blood clotting; stroke;
KW thrombosis; peripheral arterial occlusion; pulmonary embolism;
KW myocardiothrombosis; diagnosis; therapy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 124..915
FT /*tag= a
FT sig_peptide 124..184
FT /*tag= b
FT mat_peptide 185..912
FT /*tag= c
PN WO9854199-A1.
PD 03-DEC-1998.
PF 27-MAY-1998; U10728.
PR 28-MAY-1997; US-048000.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Ebner R, Moore PA, Ruben SM;
DR WPI; 99-070207/06.
DR P-PSDB; W87769.
PT New tissue plasminogen activator-like protease - useful in the
PT diagnosis and treatment of circulatory system-related disorders
PS Claim 2; Page 54-56; 76pp; English.
CC This DNA sequence includes a coding region for a novel human tissue
CC plasminogen activator-like protease (t-PALP) polypeptide (see
CC W87769) that is a member of the serine protease family and a
CC homologue of tissue plasminogen activator (tPA, see W87770). It
CC was discovered in a cDNA library derived from activated monocytes.

```
CC   The 2.5 kb t-PALP message has also been detected in heart, brain,
CC   lung, placenta, liver, skeletal muscle, kidney, pancreas, spleen,
CC   thymus, prostate, testis, ovary, small intestine, colon and
CC   peripheral blood leukocytes.  t-PALP cDNA that encodes mature
CC   t-PALP polypeptide is deposited as ATCC 209023.  Isolated nucleic
CC   acids encoding amino acids -21 to 242, -20 to 242, 1-242, 4-63
CC   (kringle domain) and 64-242 (protease domain) of t-PALP, or
CC   encoding epitope-bearing portions of t-PALP, are also claimed as
CC   are recombinant vectors, host cells, methods for producing t-PALP
CC   polypeptides, and related nucleic acid molecules (see V99637-43).
CC   t-PALP may be used to detect and treat disorders related to the
CC   circulatory system, and to identify agonists and antagonists of
CC   t-PALP activity.  The homology between t-PALP and tPA indicates
CC   that t-PALP may be involved in the regulation of normal and
CC   abnormal clotting in e.g. stroke, deep-vein thrombosis, peripheral
CC   arterial occlusion, pulmonary embolism and myocardiothrombosis.
SQ   Sequence   2329 BP;    548 A;    605 C;    676 G;    500 T;

  Query Match          100.0%;  Score 2329;  DB 60;  Length 2329;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches  2329;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 ttaccagaacagcataacaagggcaggtctgactgcaagctgggactgggaggcagagcc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 TTACCAGAACAGCATAACAAGGGCAGGTCTGACTGCAAGCTGGGACTGGGAGGCAGAGCC 60

Db     61 gccgccaaggggggcctcggttaaacactggtcgttcaatcacctgcaagacgaagaggca 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 GCCGCCAAGGGGGCCTCGGTTAAACACTGGTCGTTCAATCACCTGCAAGACGAAGAGGCA 120

Db    121 aggatgctgttggcctgggtacaagcattcctcgtcagcaacatgctcctagcagaagcc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121 AGGATGCTGTTGGCCTGGGTACAAGCATTCCTCGTCAGCAACATGCTCCTAGCAGAAGCC 180

Db    181 tatggatctggaggctgtttctgggacaacggccacctgtaccgggaggaccagacctcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    181 TATGGATCTGGAGGCTGTTTCTGGGACAACGGCCACCTGTACCGGGAGGACCAGACCTCC 240

Db    241 cccgcgccgggcctccgctgcctcaactggctggacgcgcagagcgggctggcctcggcc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    241 CCCGCGCCGGGCCTCCGCTGCCTCAACTGGCTGGACGCGCAGAGCGGGCTGGCCTCGGCC 300

Db    301 cccgtgtcgggggccggcaatcacagttactgccgaaacccggacgaggacccgcgcggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    301 CCCGTGTCGGGGGCCGGCAATCACAGTTACTGCCGAAACCCGGACGAGGACCCGCGCGGG 360

Db    361 ccctggtgctacgtcagtggcgaggccggcgtccctgagaaacggccttgcgaggacctg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    361 CCCTGGTGCTACGTCAGTGGCGAGGCCGGCGTCCCTGAGAAACGGCCTTGCGAGGACCTG 420

Db    421 cgctgtccagagaccacctcccaggccctgccagccttcacgacagaaatccaggaagcg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    421 CGCTGTCCAGAGACCACCTCCCAGGCCCTGCCAGCCTTCACGACAGAAATCCAGGAAGCG 480

Db    481 tctgaagggccaggtgcagatgaggtgcaggtgttcgctcctgccaacgccctgcccgct 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    481 TCTGAAGGGCCAGGTGCAGATGAGGTGCAGGTGTTCGCTCCTGCCAACGCCCTGCCCGCT 540

Db    541 cggagtgaggcggcagctgtgcagccagtgattgggatcagccagcgggtgcggatgaac 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    541 CGGAGTGAGGCGGCAGCTGTGCAGCCAGTGATTGGGATCAGCCAGCGGGTGCGGATGAAC 600

Db    601 tccaaggagaaaaaggacctgggaactctgggctacgtgctgggcattaccatgatggtg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    601 TCCAAGGAGAAAAAGGACCTGGGAACTCTGGGCTACGTGCTGGGCATTACCATGATGGTG 660

Db    661 atcatcattgccatcggagctggcatcatcttgggctactcctacaagaggggaaggat 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    661 ATCATCATTGCCATCGGAGCTGGCATCATCTTGGGCTACTCCTACAAGAGGGGGAAGGAT 720

Db    721 ttgaaagaacagcatgatcagaaagtatgtgagagggagatgcagcgaatcactctgccc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    721 TTGAAAGAACAGCATGATCAGAAAGTATGTGAGAGGGAGATGCAGCGAATCACTCTGCCC 780

Db    781 ttgtctgccttcaccaaccccacctgtgagattgtggatgagaagactgtcgtggtccac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    781 TTGTCTGCCTTCACCAACCCCACCTGTGAGATTGTGGATGAGAAGACTGTCGTGGTCCAC 840

Db    841 accagccagactccagttgaccctcaggagggcagcacccccctatgggccaggccggg 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    841 ACCAGCCAGACTCCAGTTGACCCTCAGGAGGGCAGCACCCCCCTTATGGGCCAGGCCGGG 900

Db    901 actcctggggcctgagcccccccagtgggcaggagcccatgcagacactggtgcaggaca 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    901 ACTCCTGGGGCCTGAGCCCCCCCAGTGGGCAGGAGCCCATGCAGACACTGGTGCAGGACA 960

Db    961 gcccaccctcctacagctaggaggaactaccactttgtgttctggttaaaaccctaccac 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    961 GCCCACCCTCCTACAGCTAGGAGGAACTACCACTTTGTGTTCTGGTTAAAACCCTACCAC 1020

Db   1021 tcccccgctttttttggcgaatcctagtaagagtgacagaagcaggtggccctgtgggctg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1021 TCCCCCGCTTTTTTGGCGAATCCTAGTAAGAGTGACAGAAGCAGGTGGCCCTGTGGGCTG 1080

Db   1081 agggtaaggctgggtagggtcctaacagtgctccttgtccatcccttggagcagattttg 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1081 AGGGTAAGGCTGGGTAGGGTCCTAACAGTGCTCCTTGTCCATCCCTTGGAGCAGATTTTG 1140

Db   1141 tctgtggatggagacagtggcagctcccacagtgatgctgctgctaagggcttccaaaca 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1141 TCTGTGGATGGAGACAGTGGCAGCTCCCACAGTGATGCTGCTGCTAAGGGCTTCCAAACA 1200

Db   1201 ttgcctgcacccctggaactgaaccagggatagacggggagctcccccaggctcctctgt 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1201 TTGCCTGCACCCCTGGAACTGAACCAGGGATAGACGGGGAGCTCCCCCAGGCTCCTCTGT 1260

Db   1261 gctttactaagatggctcagtctccactgtgggcttgagtggcatacactgttattcatg 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1261 GCTTTACTAAGATGGCTCAGTCTCCACTGTGGGCTTGAGTGGCATACACTGTTATTCATG 1320

Db   1321 gttaaggtaaagcaggtcaagggatggcattgaaaaaatatatttagtttttaaaatatt 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1321 GTTAAGGTAAAGCAGGTCAAGGGATGGCATTGAAAAAATATATTTAGTTTTTAAAATATT 1380

Db   1381 tgggatggaactccctactgacctctgacaactggaaacgagtttgtactgaagtcagaa 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1381 TGGGATGGAACTCCCTACTGACCTCTGACAACTGGAAACGAGTTTGTACTGAAGTCAGAA 1440

Db   1441 ctttgggttgggaatgagatctaggttgtggctgctggtatgcttcagcttgctggcaat 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1441 CTTTGGGTTGGGAATGAGATCTAGGTTGTGGCTGCTGGTATGCTTCAGCTTGCTGGCAAT 1500

Db   1501 gatgtgccttgacaaccgtgggccaggcctgggcccagggactcttcctgtttcataagg 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1501 GATGTGCCTTGACAACCGTGGGCCAGGCCTGGGCCCAGGGACTCTTCCTGTTTCATAAGG 1560

Db   1561 aaaggaagaattgcactgagcattccacttaggaagaggatagagaaggatctgctccgc 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1561 AAAGGAAGAATTGCACTGAGCATTCCACTTAGGAAGAGGATAGAGAAGGATCTGCTCCGC 1620

Db   1621 ctttggccacaggagcagaggcagacctgggatgccccagtttctcttcagggatggata 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1621 CTTTGGCCACAGGAGCAGAGGCAGACCTGGGATGCCCCAGTTTCTCTTCAGGGATGGATA 1680

Db   1681 gtgacctgtcttcattttgcacaggtaagagagtagttagctaacctatgggaattatac 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1681 GTGACCTGTCTTCATTTTGCACAGGTAAGAGAGTAGTTAGCTAACCTATGGGAATTATAC 1740

Db   1741 tgtggggccttgtgagctgcttctaagaggctaacctggaaactaagctcagaggcaagg 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1741 TGTGGGGCCTTGTGAGCTGCTTCTAAGAGGCTAACCTGGAAACTAAGCTCAGAGGCAAGG 1800

Db   1801 taataaagcacttcagggcttgctccccaagtgggcctgatttagcaggtggtctgcggg 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1801 TAATAAAGCACTTCAGGGCTTGCTCCCCAAGTGGGCCTGATTTAGCAGGTGGTCTGCGGG 1860
```

```
Db    1861 cgtccaggtcagcaccttcctgtagggcactggggctagggtcacagcccctaactcata 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1861 CGTCCAGGTCAGCACCTTCCTGTAGGGCACTGGGGCTAGGGTCACAGCCCCTAACTCATA 1920

Db    1921 aagcaatcaaagaaccattagaaagggctcattaagccttttggacacaggaccccagag 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1921 AAGCAATCAAAGAACCATTAGAAAGGGCTCATTAAGCCTTTTGGACACAGGACCCCAGAG 1980

Db    1981 aggaaaaagtgacttgcccaaggtcgtaagcaagctactggcatggcaagagcccagctt 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1981 AGGAAAAAGTGACTTGCCCAAGGTCGTAAGCAAGCTACTGGCATGGCAAGAGCCCAGCTT 2040

Db    2041 cctgacggagcgcaacatttctccactgcactgtgctagcagctcagcagggcctctaac 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2041 CCTGACGGAGCGCAACATTTCTCCACTGCACTGTGCTAGCAGCTCAGCAGGGCCTCTAAC 2100

Db    2101 ctgtgatgtcacactcaagaggccttggcagctcctagccatagagcttcctttccagaa 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2101 CTGTGATGTCACACTCAAGAGGCCTTGGCAGCTCCTAGCCATAGAGCTTCCTTTCCAGAA 2160

Db    2161 cccttccactgcccaatgtggagacaggggttagtggggctttctatggagccatctgct 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2161 CCCTTCCACTGCCCAATGTGGAGACAGGGGTTAGTGGGGCTTTCTATGGAGCCATCTGCT 2220

Db    2221 ttggggacctagacctcaggtggtctcttggtgttagtgatgctggagaagagaatatta 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2221 TTGGGGACCTAGACCTCAGGTGGTCTCTTGGTGTTAGTGATGCTGGAGAAGAGAATATTA 2280

Db    2281 ctggtttctacttttctataaaggcatttctctataaaaaaaaaaaaaa 2329
           |||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2281 CTGGTTTCTACTTTTCTATAAAGGCATTTCTCTATAAAAAAAAAAAAA 2329
```

RESULT 2
ID V99640 standard; DNA; 399 BP.
AC V99640;
DT 29-MAR-1999 (first entry)
DE Human t-PALP-related DNA clone HLMFG34R.
KW Tissue plasminogen activator-like protease; t-PALP; human;
KW circulatory system-related disorder; blood clotting; stroke;
KW thrombosis; peripheral arterial occlusion; pulmonary embolism;
KW myocardiothrombosis; diagnosis; therapy; ss.
OS Homo sapiens.
PN WO9854199-A1.
PD 03-DEC-1998.
PF 27-MAY-1998; U10728.
PR 28-MAY-1997; US-048000.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Ebner R, Moore PA, Ruben SM;
DR WPI; 99-070207/06.
PT New tissue plasminogen activator-like protease - useful in the
PT diagnosis and treatment of circulatory system-related disorders
PS Claim 20; Page 60; 76pp; English.
CC This DNA sequence shows homology to DNA (see V99636) coding for
CC novel human tissue plasminogen activator-like protease (t-PALP, see
CC W87769). It was obtained from cDNA clone HLMFG34R. A nucleic acid
CC molecule comprising a polynucleotide having a sequence at least
CC 95% identical to one of 7 t-PALP related sequences (see V99637-43)
CC is claimed. Full-length t-PALP DNA (also claimed) was discovered
CC in a cDNA library derived from activated monocytes. The 2.5 kb
CC t-PALP message has also been detected in heart, brain, lung,
CC placenta, liver, skeletal muscle, kidney, pancreas, spleen, thymus,
CC prostate, testis, ovary, small intestine, colon and peripheral
CC blood leukocytes. Vectors, host cells and methods for producing
CC t-PALP polypeptides are claimed. The homology between t-PALP and
CC tissue plasminogen activator indicates that t-PALP may be involved
CC in the regulation of normal and abnormal clotting in e.g. stroke,
CC deep-vein thrombosis, peripheral arterial occlusion, pulmonary
CC embolism and myocardiothrombosis.
SQ Sequence 399 BP; 102 A; 74 C; 116 G; 107 T;

Query Match 11.6%; Score 270; DB 60; Length 399;
Best Local Similarity 91.0%; Pred. No. 3.38e-159;
Matches 355; Conservative 0; Mismatches 25; Indels 10; Gaps 9;

```
Db      13 aactggaaacgagtttgtacagaagtcagaactttgggttaggaatgagatctaggttgt 72
           ||||||||||||||||||||| ||||||||||||||||||| ||||||||||||||||||
Qy    1410 AACTGGAAACGAGTTTGTACTGAAGTCAGAACTTTGGGTTGGGAATGAGATCTAGGTTGT 1469

Db      73 ggctgctggtatgcttca--ttgctggcaataatgtgccttgacaaccgtgggccaggcc 130
           |||||||||||||||||  |||||||||| |||||||||||||||||||||||||||||
Qy    1470 GGCTGCTGGTATGCTTCAGCTTGCTGGCAATGATGTGCCTTGACAACCGTGGGCCAGGCC 1529

Db     131 tgggaccagggactcttcctgtttcataaggaaaggaagaattgcactgagcattccact 190
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1530 TGGGCCCAGGGACTCTTCCTGTTTCATAAGGAAAGGAAGAATTGCACTGAGCATTCCACT 1589

Db     191 taggaagaggatagagcaaggaatctgctccgc-tttggccacaggagcagaggcagacc 249
           |||||||||||||||| ||||| |||||||||| ||||||||||||||||||||||||||
Qy    1590 TAGGAAGAGGATAGAG-AAGGA-TCTGCTCCGCCTTTGGCCACAGGAGCAGAGGCAGACC 1647

Db     250 tgggatgccccagttctctttcagggatgggatagtgacctgtcttacattttgcacagg 309
           ||||||||||||||    |||||||||| |||||||||||||| ||||||||||||||
Qy    1648 TGGGATGCCCCAGTTTCTCTTCAGGGATGG-ATAGTGACCTGTCTT-CATTTTGCACAGG 1705

Db     310 taaagagagttagttagctaacctattgggctttattacttggggcttgtgagctgcttt 369
           ||| |||||| |||||||||||||| ||  ||      | ||| ||||||||||||||
Qy    1706 TAA-GAGAGT-AGTTAGCTAACCTATGGGAATTATACTGTGGGGCCTTGTGAGCTGCTTC 1763

Db     370 ttaagaggttaacctggaactaaagttcag 399
           | |||||| ||||||||||   ||| ||||
Qy    1764 T-AAGAGGCTAACCTGGAAACTAAGCTCAG 1792
```

RESULT 3
ID V99642 standard; DNA; 472 BP.
AC V99642;
DT 29-MAR-1999 (first entry)
DE Human t-PALP-related DNA clone HSSAX27R.
KW Tissue plasminogen activator-like protease; t-PALP; human;
KW circulatory system-related disorder; blood clotting; stroke;
KW thrombosis; peripheral arterial occlusion; pulmonary embolism;
KW myocardiothrombosis; diagnosis; therapy; ss.
OS Homo sapiens.
PN WO9854199-A1.
PD 03-DEC-1998.
PF 27-MAY-1998; U10728.
PR 28-MAY-1997; US-048000.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Ebner R, Moore PA, Ruben SM;
DR WPI; 99-070207/06.
PT New tissue plasminogen activator-like protease - useful in the
PT diagnosis and treatment of circulatory system-related disorders
PS Claim 20; Page 61; 76pp; English.
CC This DNA sequence shows homology to DNA (see V99636) coding for
CC novel human tissue plasminogen activator-like protease (t-PALP, see
CC W87769). It was obtained from cDNA clone HSSAX27R. A nucleic acid
CC molecule comprising a polynucleotide having a sequence at least
CC 95% identical to one of 7 t-PALP related sequences (see V99637-43)
CC is claimed. Full-length t-PALP DNA (also claimed) was discovered
CC in a cDNA library derived from activated monocytes. The 2.5 kb
CC t-PALP message has also been detected in heart, brain, lung,
CC placenta, liver, skeletal muscle, kidney, pancreas, spleen, thymus,
CC prostate, testis, ovary, small intestine, colon and peripheral
CC blood leukocytes. Vectors, host cells and methods for producing
CC t-PALP polypeptides are claimed. The homology between t-PALP and
CC tissue plasminogen activator indicates that t-PALP may be involved
CC in the regulation of normal and abnormal clotting in e.g. stroke,
CC deep-vein thrombosis, peripheral arterial occlusion, pulmonary
CC embolism and myocardiothrombosis.
SQ Sequence 472 BP; 100 A; 112 C; 149 G; 111 T;

Query Match 11.5%; Score 267; DB 60; Length 472;
Best Local Similarity 84.4%; Pred. No. 3.88e-157;
Matches 389; Conservative 0; Mismatches 62; Indels 10; Gaps 10;

```
Db     14 gagagggagatgcagcgaatcactctgcccttgtctgccttcaccaaccccacctgtgag 73
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    751 GAGAGGGAGATGCAGCGAATCACTCTGCCCTTGTCTGCCTTCACCAACCCCACCTGTGAG 810

Db     74 attgtggatgagaagactgtcgtggtccacaccagccagactccagttgaccctcaggag 133
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    811 ATTGTGGATGAGAAGACTGTCGTGGTCCACACCAGCCAGACTCCAGTTGACCCTCAGGAG 870

Db    134 ggcagcaccccccttatgggaccaggccggggactcctggggcctgagccccccagtggg 193
          |||||||||||||||||| |||||||||| ||||||||||||||||||||    ||
Qy    871 GGCAGCACCCCCCTTATGGG-CCAGGCCGGG-ACTCCTGGGGCCTGAGCCCCCCCAGTGG 928

Db    194 gcaggagccatggcagacactggtgcaggacagccaccctccttacagctaggggggaact 253
          |||||||||   |||||||||||||||||||||  ||  | |||||||||| ||||||
Qy    929 GCAGGAGCCCATGCAGACACTGGTGCAGGACAGCCCACCCTCCTACAGCTAGGAGGAACT 988

Db    254 accactttgtgtttctggtttaaaaccctaccactcccggatttttttggcggattcctta 313
          |||||||||||||| |||||| |||||||||||||||||    ||||| |  || |||| |
Qy    989 ACCACTTTGTGTT-CTGGTT-AAAACCCTACCACTCCCCCGCTTTTTTGGCGAATCCT-A 1045

Db    314 gttaagagtacagaagcaggtgggcctatggcttggagggtaaggtggggtagggttcct 373
          || |     |||||||||||||| ||| |||  || |||||||||  ||||||||| |||
Qy   1046 GTAAGAGTGACAGAAGCAGGTGGCCCTGTGGGCTG-AGGGTAAGGCTGGGTAGGGT-CCT 1103

Db    374 aaaagtgggttcttggttgctcctgggaggaagattttggttttggtggggacagtggca 433
          || |||| | ||||      ||| |||| |   |||   | | | ||| ||||||||||
Qy   1104 AACAGTGC-TCCTTGTCCATCCCTTGGAGCAGATTTTGTCTGTGGATGGAGACAGTGGCA 1162

Db    434 gtttccacag-gttgttg-tgttaaggggttcaaaaaattg 472
          | | |||||| | || || || |||||| ||| ||| ||||
Qy   1163 GCTCCCACAGTGATGCTGCTGCTAAGGGCTTCCAAACATTG 1203
```

RESULT 4
ID V99637 standard; DNA; 250 BP.
AC V99637;
DT 29-MAR-1999 (first entry)
DE Human t-PALP-related DNA clone HTAAM28R.
KW Tissue plasminogen activator-like protease; t-PALP; human;
KW circulatory system-related disorder; blood clotting; stroke;
KW thrombosis; peripheral arterial occlusion; pulmonary embolism;
KW myocardiothrombosis; diagnosis; therapy; ss.
OS Homo sapiens.
PN WO9854199-A1.
PD 03-DEC-1998.
PF 27-MAY-1998; U10728.
PR 28-MAY-1997; US-048000.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Ebner R, Moore PA, Ruben SM;
DR WPI; 99-070207/06.
PT New tissue plasminogen activator-like protease - useful in the
PT diagnosis and treatment of circulatory system-related disorders
PS Claim 20; Page 59; 76pp; English.
CC This DNA sequence shows homology to DNA (see V99636) coding for
CC novel human tissue plasminogen activator-like protease (t-PALP, see
CC W87769). It was obtained from cDNA clone HTAAM28R. A nucleic acid
CC molecule comprising a polynucleotide having a sequence at least
CC 95% identical to one of 7 t-PALP related sequences (see V99637-43)
CC is claimed. Full-length t-PALP DNA (also claimed) was discovered
CC in a cDNA library derived from activated monocytes. The 2.5 kb
CC t-PALP message has also been detected in heart, brain, lung,
CC placenta, liver, skeletal muscle, kidney, pancreas, spleen, thymus,
CC prostate, testis, ovary, small intestine, colon and peripheral
CC blood leukocytes. Vectors, host cells and methods for producing
CC t-PALP polypeptides are claimed. The homology between t-PALP and
CC tissue plasminogen activator indicates that t-PALP may be involved
CC in the regulation of normal and abnormal clotting in e.g. stroke,
CC deep-vein thrombosis, peripheral arterial occlusion, pulmonary
CC embolism and myocardiothrombosis.
SQ Sequence 250 BP; 67 A; 52 C; 69 G; 62 T;

Query Match 10.2%; Score 238; DB 60; Length 250;
Best Local Similarity 99.2%; Pred. No. 2.82e-137;
Matches 250; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

```
Db      1 attgcactgagcattccacttaggaagaggatagagaaggatctgctccgcctttggcca 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1570 ATTGCACTGAGCATTCCACTTAGGAAGAGGATAGAGAAGGATCTGCTCCGCCTTTGGCCA 1629

Db     61 caggagcagaggcagacctgggatgcccca-tttctcttcagggatggatagtgacctgt 119
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Qy   1630 CAGGAGCAGAGGCAGACCTGGGATGCCCCAGTTTCTCTTCAGGGATGGATAGTGACCTGT 1689

Db    120 cttcattttgcacaggtaagagagtagttagctaacctatgggaattatactgtggggcc 179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1690 CTTCATTTTGCACAGGTAAGAGAGTAGTTAGCTAACCTATGGGAATTATACTGTGGGGCC 1749

Db    180 ttgt-agctgcttctaagaggctaacctggaaactaagctcagaggcaaggtaataaagc 238
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1750 TTGTGAGCTGCTTCTAAGAGGCTAACCTGGAAACTAAGCTCAGAGGCAAGGTAATAAAGC 1809

Db    239 acttcagggctt 250
          ||||||||||||
Qy   1810 ACTTCAGGGCTT 1821
```

RESULT 5
ID V99638 standard; DNA; 247 BP.
AC V99638;
DT 29-MAR-1999 (first entry)
DE Human t-PALP-related DNA clone HFKBA12R.
KW Tissue plasminogen activator-like protease; t-PALP; human;
KW circulatory system-related disorder; blood clotting; stroke;
KW thrombosis; peripheral arterial occlusion; pulmonary embolism;
KW myocardiothrombosis; diagnosis; therapy; ss.
OS Homo sapiens.
PN WO9854199-A1.
PD 03-DEC-1998.
PF 27-MAY-1998; U10728.
PR 28-MAY-1997; US-048000.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Ebner R, Moore PA, Ruben SM;
DR WPI; 99-070207/06.
PT New tissue plasminogen activator-like protease - useful in the
PT diagnosis and treatment of circulatory system-related disorders
PS Claim 20; Page 59; 76pp; English.
CC This DNA sequence shows homology to DNA (see V99636) coding for
CC novel human tissue plasminogen activator-like protease (t-PALP, see
CC W87769). It was obtained from cDNA clone HFKBA12R. A nucleic acid
CC molecule comprising a polynucleotide having a sequence at least
CC 95% identical to one of 7 t-PALP related sequences (see V99637-43)
CC is claimed. Full-length t-PALP DNA (also claimed) was discovered
CC in a cDNA library derived from activated monocytes. The 2.5 kb
CC t-PALP message has also been detected in heart, brain, lung,
CC placenta, liver, skeletal muscle, kidney, pancreas, spleen, thymus,
CC prostate, testis, ovary, small intestine, colon and peripheral
CC blood leukocytes. Vectors, host cells and methods for producing
CC t-PALP polypeptides are claimed. The homology between t-PALP and
CC tissue plasminogen activator indicates that t-PALP may be involved
CC in the regulation of normal and abnormal clotting in e.g. stroke,
CC deep-vein thrombosis, peripheral arterial occlusion, pulmonary
CC embolism and myocardiothrombosis.
SQ Sequence 247 BP; 73 A; 63 C; 60 G; 51 T;

Query Match 10.1%; Score 235; DB 60; Length 247;
Best Local Similarity 99.2%; Pred. No. 3.16e-135;
Matches 247; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

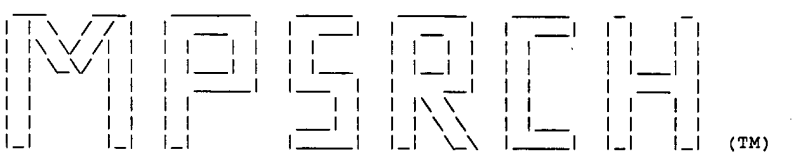
```
Db      1 atagagaaatgcctttatagaaaagtagaaaccagtaatattctcttctccagcatcact 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2315 ATAGAGAAATGCCTTTATAGAAAAGTAGAAACCAGTAATATTCTCTTCTCCAGCATCACT 2256

Db     61 aacaccaagagaccacctgaggtctaggtccccaaagcagatggctccatagaaagcccc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2255 AACACCAAGAGACCACCTGAGGTCTAGGTCCCCAAAGCAGATGGCTCCATAGAAAGCCCC 2196
```

```
Db    121 actaaccc--gtctccacattgggcagtggaagggttctggaaaggaagctctatggcta 178
          ||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2195 ACTAACCCCTGTCTCCACATTGGGCAGTGGAAGGGTTCTGGAAAGGAAGCTCTATGGCTA 2136

Db    179 ggagctgccaaggcctcttgagtgtgacatcacaggttagaggccctgctgagctgctag 238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   2135 GGAGCTGCCAAGGCCTCTTGAGTGTGACATCACAGGTTAGAGGCCCTGCTGAGCTGCTAG 2076

Db    239 cacagtgca 247
          |||||||||
Cp   2075 CACAGTGCA 2067


RESULT    6
ID   V99639 standard; DNA; 461 BP.
AC   V99639;
DT   29-MAR-1999  (first entry)
DE   Human t-PALP-related DNA clone HAPBL24R.
KW   Tissue plasminogen activator-like protease; t-PALP; human;
KW   circulatory system-related disorder; blood clotting; stroke;
KW   thrombosis; peripheral arterial occlusion; pulmonary embolism;
KW   myocardiothrombosis; diagnosis; therapy; ss.
OS   Homo sapiens.
PN   WO9854199-A1.
PD   03-DEC-1998.
PF   27-MAY-1998; U10728.
PR   28-MAY-1997; US-048000.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Ebner R, Moore PA, Ruben SM;
DR   WPI; 99-070207/06.
PT   New tissue plasminogen activator-like protease - useful in the
PT   diagnosis and treatment of circulatory system-related disorders
PS   Claim 20; Page 59-60; 76pp; English.
CC   This DNA sequence shows homology to DNA (see V99636) coding for
CC   novel human tissue plasminogen activator-like protease (t-PALP, see
CC   W87769).  It was obtained from cDNA clone HAPBL24R.  A nucleic acid
CC   molecule comprising a polynucleotide having a sequence at least
CC   95% identical to one of 7 t-PALP related sequences (see V99637-43)
CC   is claimed.  Full-length t-PALP DNA (also claimed) was discovered
CC   in a cDNA library derived from activated monocytes.  The 2.5 kb
CC   t-PALP message has also been detected in heart, brain, lung,
CC   placenta, liver, skeletal muscle, kidney, pancreas, spleen, thymus,
CC   prostate, testis, ovary, small intestine, colon and peripheral
CC   blood leukocytes.  Vectors, host cells and methods for producing
CC   t-PALP polypeptides are claimed.  The homology between t-PALP and
CC   tissue plasminogen activator indicates that t-PALP may be involved
CC   in the regulation of normal and abnormal clotting in e.g. stroke,
CC   deep-vein thrombosis, peripheral arterial occlusion, pulmonary
CC   embolism and myocardiothrombosis.
SQ   Sequence   461 BP;    95 A;    108 C;    145 G;    113 T;

  Query Match            8.2%;  Score 191;  DB 60;  Length 461;
  Best Local Similarity  92.3%;  Pred. No. 2.63e-105;
  Matches   263;  Conservative    0;  Mismatches  12;  Indels  10;  Gaps   9;

Db     15 aacagcataacaagggtaggtctgactgca-gctgggactgggaggcagagca--cgcca 71
          ||||||||||||||| |||||||||||| ||||||||||||||||||||   |||||
Qy      8 AACAGCATAACAAGGGCAGGTCTGACTGCAAGCTGGGACTGGGAGGCAGAGCCGCCGCCA 67

Db     72 agggggcctcggttaaacactggtcgttcaatcacctgcaa-acgaggaggcaaggatgc 130
          |||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||||
Qy     68 AGGGGGCCTCGGTTAAACACTGGTCGTTCAATCACCTGCAAGACGAAGAGGCAAGGATGC 127

Db    131 tgttggcctgggtaca-gcattcctggtcagcaacatgctcctagcgtaagcctatggat 189
          |||||||||||||||| |||||||| ||||||||||||||||||  ||||||||||||
Qy    128 TGTTGGCCTGGGTACAAGCATTCCTCGTCAGCAACATGCTCCTAGCAGAAGCCTATGGAT 187

Db    190 ctggaggctgtttctgggacaacggccac-tgtacccggaggaccagaccttccc-g-gc 246
          |||||||||||||||||||||||||||||| |||||| ||||||||||||| ||| | ||
Qy    188 CTGGAGGCTGTTTCTGGGACAACGGCCACCTGTACCGGGAGGACCAGACCTCCCCCGCGC 247

Db    247 cggtcctc-g-tgcctcaactggctggacgcgcagggctgcctgg 289
          ||| |||| | ||||||||||||||||||||||||| || | ||||
Qy    248 CGGGCCTCCGCTGCCTCAACTGGCTGGACGCGCAGAGCGGGCTGG 292


RESULT    7
ID   V99641 standard; DNA; 334 BP.
AC   V99641;
DT   29-MAR-1999  (first entry)
DE   Human t-PALP-related DNA clone HHPGT42R.
KW   Tissue plasminogen activator-like protease; t-PALP; human;
KW   circulatory system-related disorder; blood clotting; stroke;
KW   thrombosis; peripheral arterial occlusion; pulmonary embolism;
KW   myocardiothrombosis; diagnosis; therapy; ss.
OS   Homo sapiens.
PN   WO9854199-A1.
PD   03-DEC-1998.
PF   27-MAY-1998; U10728.
PR   28-MAY-1997; US-048000.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Ebner R, Moore PA, Ruben SM;
DR   WPI; 99-070207/06.
PT   New tissue plasminogen activator-like protease - useful in the
PT   diagnosis and treatment of circulatory system-related disorders
PS   Claim 20; Page 61; 76pp; English.
CC   This DNA sequence shows homology to DNA (see V99636) coding for
CC   novel human tissue plasminogen activator-like protease (t-PALP, see
CC   W87769).  It was obtained from cDNA clone HHPGT42R.  A nucleic acid
CC   molecule comprising a polynucleotide having a sequence at least
CC   95% identical to one of 7 t-PALP related sequences (see V99637-43)
CC   is claimed.  Full-length t-PALP DNA (also claimed) was discovered
CC   in a cDNA library derived from activated monocytes.  The 2.5 kb
CC   t-PALP message has also been detected in heart, brain, lung,
CC   placenta, liver, skeletal muscle, kidney, pancreas, spleen, thymus,
CC   prostate, testis, ovary, small intestine, colon and peripheral
CC   blood leukocytes.  Vectors, host cells and methods for producing
CC   t-PALP polypeptides are claimed.  The homology between t-PALP and
CC   tissue plasminogen activator indicates that t-PALP may be involved
CC   in the regulation of normal and abnormal clotting in e.g. stroke,
CC   deep-vein thrombosis, peripheral arterial occlusion, pulmonary
CC   embolism and myocardiothrombosis.
SQ   Sequence   334 BP;    74 A;    82 C;    106 G;    72 T;

  Query Match            6.4%;  Score 150;  DB 60;  Length 334;
  Best Local Similarity  88.0%;  Pred. No. 9.67e-78;
  Matches   249;  Conservative    0;  Mismatches  21;  Indels  13;  Gaps  10;

Db     13 ggacaggtctgactgca-gctgggactgggaggcagagccgtc---aaggggcctcggt 68
          || |||||||||||||| |||||||||||||||||||||| |   ||||||||||||
Qy     21 GGGCAGGTCTGACTGCAAGCTGGGACTGGGAGGCAGAGCCGCCGCCAAGGGGCCTCGGT 80

Db     69 taaacactggtcgttcaatcacctgcaa--cga-gaggcaaggatgctgttggcctgggt 125
          |||||||||||||||||||||||||||||  ||| |||||||||||||||||||||||||
Qy     81 TAAACACTGGTCGTTCAATCACCTGCAAGACGAAGAGGCAAGGATGCTGTTGGCCTGGGT 140

Db    126 acaagcattcct-gtcagcaacatgctcctagcagaaagcctatggatctgggaggctgt 184
          |||||||||||| |||||||||||||||||||||||| |||||||||||||| |||||||
Qy    141 ACAAGCATTCCTCGTCAGCAACATGCTCCTAGCAGAA-GCCTATGGATCTGG-AGGCTGT 198

Db    185 ttctgggacaacggccacctgtaccgg-aggaccagacctccccggc-cgggccttccgt 242
          |||||||||||||||||||||||||||| |||||||||||||||| || | || | ||||
Qy    199 TTCTGGGACAACGGCCACCTGTACCGGGAGGACCAGACCTCCCCCGCGCCGGGCCTCCGC 258

Db    243 ggccttcaattggtttgacgtggcaaaggggcttgtctggccc 285
          |||| ||| ||| | |||| |   | ||||| | || | ||
Qy    259 TGCCT-CAACTGGCTGGACGCGCAGAGCGGGCTGGCCTCGGCC 300


RESULT    8
ID   V99643 standard; DNA; 291 BP.
AC   V99643;
DT   29-MAR-1999  (first entry)
DE   Human t-PALP-related DNA clone HSSES93R.
KW   Tissue plasminogen activator-like protease; t-PALP; human;
KW   circulatory system-related disorder; blood clotting; stroke;
```

```
KW   thrombosis; peripheral arterial occlusion; pulmonary embolism;
KW   myocardiothrombosis; diagnosis; therapy; ss.
OS   Homo sapiens.
PN   WO9854199-A1.
PD   03-DEC-1998.
PF   27-MAY-1998; U10728.
PR   28-MAY-1997; US-048000.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Ebner R, Moore PA, Ruben SM;
DR   WPI; 99-070207/06.
PT   New tissue plasminogen activator-like protease - useful in the
PT   diagnosis and treatment of circulatory system-related disorders
PS   Claim 20; Page 62; 76pp; English.
CC   This DNA sequence shows homology to DNA (see V99636) coding for
CC   novel human tissue plasminogen activator-like protease (t-PALP, see
CC   W87769). It was obtained from cDNA clone HSSES93R. A nucleic acid
CC   molecule comprising a polynucleotide having a sequence at least
CC   95% identical to one of 7 t-PALP related sequences (see V99637-43)
CC   is claimed. Full-length t-PALP DNA (also claimed) was discovered
CC   in a cDNA library derived from activated monocytes. The 2.5 kb
CC   t-PALP message has also been detected in heart, brain, lung,
CC   placenta, liver, skeletal muscle, kidney, pancreas, spleen, thymus,
CC   prostate, testis, ovary, small intestine, colon and peripheral
CC   blood leukocytes. Vectors, host cells and methods for producing
CC   t-PALP polypeptides are claimed. The homology between t-PALP and
CC   tissue plasminogen activator indicates that t-PALP may be involved
CC   in the regulation of normal and abnormal clotting in e.g. stroke,
CC   deep-vein thrombosis, peripheral arterial occlusion, pulmonary
CC   embolism and myocardiothrombosis.
SQ   Sequence   291 BP;    66 A;    67 C;    85 G;    73 T;

  Query Match             3.2%;  Score 74;  DB 60;  Length 291;
  Best Local Similarity  98.8%;  Pred. No. 2.54e-28;
  Matches    80;  Conservative    0;  Mismatches   0;  Indels   1;  Gaps   1;

Db    151 ggctacgtgctgggcattaccatgatggtgatcatcattgccatcggagctggcatcatc 210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    631 GGCTACGTGCTGGGCATTACCATGATGGTGATCATCATTGCCATCGGAGCTGGCATCATC 690

Db    211 ttgggctactc-tacaagagg 230
          ||||||||||| |||||||||
Qy    691 TTGGGCTACTCCTACAAGAGG 711


RESULT    9
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag=  a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;    21 A;    47 C;    17 G;    11 T;    108 Others;

  Query Match             1.9%;  Score 44;  DB 1;  Length 204;
  Best Local Similarity  10.7%;  Pred. No. 2.32e-10;
  Matches    11;  Conservative   58;  Mismatches  33;  Indels   1;  Gaps   1;

Db     85 gymrttthhyrrmrbnvyrdynrsdaaa-wyccyrrsvkydccynachhddhyvybbbvy 143
          |:::  |:::::::: ::::  ::  |  ::||:::::: :| : ||:::::::: ::::
Qy   2010 GCAAGCTACTGGCATGGCAAGAGCCCAGCTTCCTGACGGAGCGCAACATTTCTCCACTGC 2069

Db    144 nvhnhnncncccbnnhvchnvhbnnhrnwayvrhdarrddvhc 186
            :: :    | :  :: : : :  :  :|::  : : ::::|
Qy   2070 ACTGTGCTAGCAGCTCAGCAGGGCCTCTAACCTGTGATGTCAC 2112


RESULT   10
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp. The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match             1.8%;  Score 42;  DB 9;  Length 91;
  Best Local Similarity   8.9%;  Pred. No. 3.07e-09;
  Matches     5;  Conservative   44;  Mismatches   7;  Indels   0;  Gaps   0;

Db      7 ggcgssvhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsvct 62
          ||| ::::::::  :::::::: ::::::::: :::::::::::::: :: :::||
Qy    894 GGCCGGGACTCCTGGGGCCTGAGCCCCCCCAGTGGGCAGGAGCCCATGCAGACACT 949


RESULT   11
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
```

```
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match            1.8%;  Score 43;  DB 9;  Length 91;
  Best Local Similarity  11.9%;  Pred. No. 8.48e-10;
  Matches    7;  Conservative   44;  Mismatches   8;  Indels   0;  Gaps   0;

Db      7 ggcgssvhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsvctcaa 65
          ||| : : :::: :::: ::::::: : :::::::::::::::::::::::::::| |||
Cp   1524 GGCCCACGGTTGTCAAGGCACATCATTGCCAGCAAGCTGAAGCATACCAGCAGCCACAA 1466


RESULT   12
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag=  a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match            1.7%;  Score 40;  DB 1;  Length 204;
  Best Local Similarity  9.4%;  Pred. No. 3.94e-08;
  Matches   11;  Conservative   63;  Mismatches  42;  Indels   1;  Gaps   1;

Db     75 yycdchvgccgymrttthhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhd- 133
          ::|:|:::|   ::   | :::::::: :: :  : :   :: |::: ::::| :   :::
Cp   1982 CTCTCTGGGGTCCTGTGTCCAAAAGGCTTAATGAGCCCTTTCTAATGGTTCTTTGATTGC 1923

Db    134 dhyvybbbvynvhnhnncncccbnnhvchnvhbnnhrnwayvrhdarrddvhccvch 190
          :::::: : : :         |:  ::|: :::  :  :|::: :|::::::  :|:
Cp   1922 TTTATGAGTTAGGGGCTGTGACCCTAGCCCCAGTGCCCTACAGGAAGGTGCTGACCT 1866


RESULT   13
ID   Q70468 standard; DNA; 114 BP.
AC   Q70468;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65154.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70468 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed activity
CC   allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match            1.6%;  Score 38;  DB 12;  Length 114;
  Best Local Similarity  4.5%;  Pred. No. 4.87e-07;
  Matches    5;  Conservative   34;  Mismatches  73;  Indels   0;  Gaps   0;

Db      3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnnn 62
          :  :  :  :  :  :  :  :  :  :  :  :|||  :  :  :  :  :  :  :
Cp    312 CCCCGACACGGGGGCCGAGGCCAGCCCGCTCTGCGCGTCCAGCCAGTTGAGGCAGCGGAG 253

Db     63 bnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
          :  :  :  :  :  :  :||   :  :  :  :  :  :  :  :  :  :
Cp    252 GCCCGGCGCGGGGGAGGTCTGGTCCTCCCGGTACAGGTGGCCGTTGTCCCAG 201


RESULT   14
ID   Q70469 standard; DNA; 114 BP.
AC   Q70469;
DT   07-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
```

```
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6,9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70469 is a generic DNA sequence used to generate random TSAR peptide
CC   This generic formula can be represented  as follows: X(TGC)(NNB)10-
CC   (TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking  restriction
CC   sites (X is not the same as Y) that are not specified further. This
CC   sequence generates peptides that are cloverleaf in structure. Other
CC   generic sequences are shown in Q70465-68. Other specific peptides
CC   generated by these generic sequences are shown in R65150-54. TSARs are
CC   concatenated heterofunctional proteins or peptides, comprising at least
CC   two functional regions - a binding domain with affinity for a ligand and
CC   a second effector peptide portion that is chemically or biologically
CC   active.They may further comprise a linker peptide between the 2 domains.
CC   The oligonucleotides are also designed so that the expressed peptide
CC   contains 2 or 4 cysteine residues positioned in, or flanking, the
CC   unpredicted or variant residues. These residues confer some degree of
CC   conformational rigidity to the peptides. The TSARs or compsns. comprising
CC   a TSAR binding domain can be used in vivo to deliver a chemically or
CC   biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
CC   or enzyme, to the specific target or on the cell. They can also replace
CC   the function of macromolecules, eg. monoclonal or polyclonal antibodies
CC   and therefore circumvent the need for complex methods of hybridoma
CC   formation or in vivo antibody production. The TSARs are easily
CC   characterised and have designed activity allowing direct and rapid
CC   detection in a screening process.
SQ   Sequence   114 BP;    0 A;    4 C;    4 G;    4 T;

  Query Match             1.6%;  Score 37;  DB 12;  Length 114;
  Best Local Similarity   5.5%;  Pred. No. 1.69e-06;
  Matches     6;  Conservative   32;  Mismatches  71;  Indels   0;  Gaps   0;

Db      3 cnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnnn 62
          |  :  :  :  :  :  :  :  :  :  :|||  :  :  :  :  :  :
Cp    312 CCCCGACACGGGGGCCGAGGCCAGCCCGCTCTGCGCGTCCAGCCAGTTGAGGCAGCGGAG 253

Db     63 bnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 111
          :  : ||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Cp    252 GCCCGGCGCGGGGGAGGTCTGGTCCTCCCGGTACAGGTGGCCGTTGTCC 204


RESULT   15
ID   Q70465 standard; DNA; 114 BP.
AC   Q70465;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65150 and R65151.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)6(TGC)(NNB)11Z(NNB)14(TGC)(NNB)3Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match             1.5%;  Score 36;  DB 12;  Length 114;
  Best Local Similarity   4.5%;  Pred. No. 5.78e-06;
  Matches     5;  Conservative   33;  Mismatches  74;  Indels   0;  Gaps   0;

Db      3 bnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
          :  :  :  :  :  :||  :  :  :  :  :  :  :  :  :  :  :
Qy    267 CTGGCTGGACGCGCAGAGCGGGCTGGCCTCGGCCCCCGTGTCGGGGGCCGGCAATCACAG 326

Db     63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnb 114
          :  :  :     :  :  :  :  :  :  :  :  :  :|||  :  :  :
Qy    327 TTACTGCCGAAACCCGGACGAGGACCCGCGCGGGCCCTGGTGCTACGTCAGT 378


Search completed: Mon Sep 20 20:55:28 1999
Job time : 1033 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 17 18:23:46 1999; MasPar time 11.68 Seconds
478.708 Million cell updates/sec
Tabular output not generated.

Title: >US-09-084-491A-2
Description: (1-263) from US09084491A.pep
Perfect Score: 1883
Sequence: 1 MLLAWVQAFLVSNMLLAEAY.....PVDPQEGSTPLMGQAGTPGA 263

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 33.039; Variance 139.441; scale 0.237

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1883	100.0	263	39	W87769	Human tissue plasmino	1.04e-174
2	256	13.6	39	36	W72641	Nervous glia cell gro	3.64e-13
3	255	13.5	39	36	W72640	Nervous glia cell gro	4.50e-13
4	184	9.8	812	15	R83959	Complete mouse plasmi	1.03e-06
5	184	9.8	812	21	W07585	Murine plasminogen se	1.03e-06
6	181	9.6	655	17	R89197	Human hepatocellular	1.88e-06
7	181	9.6	655	10	R53962	Hepatocyte growth fac	1.88e-06
8	177	9.4	701	10	R66600	Mouse L5/3 tumour sup	4.17e-06
9	177	9.4	701	22	W14271	Mouse growth factor L	4.17e-06
10	177	9.4	716	10	R66601	Mouse L5/3 tumour sup	4.17e-06
11	177	9.4	716	22	W14272	Mouse growth factor L	4.17e-06
12	176	9.3	217	28	W46423	Bovine macrophage sti	5.09e-06
13	175	9.3	380	1	R05433	cPA-P2 Hybrid plasmin	6.21e-06
14	171	9.1	504	6	R32710	Haematopoietic stem c	1.38e-05
15	170	9.0	728	4	R21976	Human Hepatocyte grow	1.68e-05
16	167	8.9	723	3	R15624	Human leukocyte-deriv	3.03e-05
17	167	8.9	723	6	R29819	TCF-II.	3.03e-05
18	167	8.9	723	15	R82685	Tumour cytotoxic fact	3.03e-05
19	167	8.9	723	2	R07144	Tumour cytotoxic fact	3.03e-05
20	167	8.9	723	36	W59923	Human leukocyte-deriv	3.03e-05
21	167	8.9	723	37	W76690	Human plasminogen-lik	3.03e-05
22	167	8.9	723	11	R57026	Human wild-type tumor	3.03e-05
23	167	8.9	727	2	R10656	Hepatic parenchymal c	3.03e-05
24	167	8.9	728	39	W88529	Human hepatocyte grow	3.03e-05
25	167	8.9	728	39	W88532	Human hepatocyte grow	3.03e-05
26	167	8.9	728	39	W88530	Human hepatocyte grow	3.03e-05
27	167	8.9	728	39	W88531	Human hepatocyte grow	3.03e-05
28	167	8.9	728	5	R25160	Human HGF.	3.03e-05
29	167	8.9	728	32	W58696	Human hepatocyte grow	3.03e-05
30	167	8.9	728	31	W42998	Recombinant human hep	3.03e-05
31	167	8.9	728	21	W00338	Human hepatic parench	3.03e-05
32	167	8.9	728	8	R52943	Human hepatocyte grow	3.03e-05
33	167	8.9	728	16	R87525	Mutant hepatocyte gro	3.03e-05
34	167	8.9	728	8	R52940	Human hepatocyte grow	3.03e-05
35	167	8.9	728	8	R52948	Human hepatocyte grow	3.03e-05
36	167	8.9	728	28	W39207	Human hepatocyte grow	3.03e-05
37	167	8.9	728	8	R52945	Human hepatocyte grow	3.03e-05
38	167	8.9	728	8	R52949	Human hepatocyte grow	3.03e-05
39	167	8.9	728	36	W59922	Human leukocyte-deriv	3.03e-05
40	167	8.9	728	5	R25676	Recombinant human hep	3.03e-05
41	167	8.9	728	8	R52942	Human hepatocyte grow	3.03e-05
42	167	8.9	728	8	R52944	Human hepatocyte grow	3.03e-05
43	167	8.9	728	8	R40862	Competative inhibitor	3.03e-05
44	167	8.9	728	19	W00340	Wild type hepatocyte	3.03e-05
45	167	8.9	728	3	R15623	Human leukocyte-deriv	3.03e-05

ALIGNMENTS

RESULT 1
ID W87769 standard; Protein; 263 AA.
AC W87769;
DT 29-MAR-1999 (first entry)
DE Human tissue plasminogen activator-like protease t-PALP.
KW Tissue plasminogen activator-like protease; t-PALP; human;
KW circulatory system-related disorder; blood clotting; stroke;
KW thrombosis; peripheral arterial occlusion; pulmonary embolism;
KW myocardiothrombosis; diagnosis; therapy.
OS Homo sapiens.

FH	Key	Location/Qualifiers
FT	Peptide	1..21
FT		/label= Sig_peptide
FT	Protein	22..263
FT		/label= Mat_protein
FT	Domain	25..84
FT		/note= "kringle domain"
FT	Domain	85..263
FT		/note= "protease domain"
FT	Peptide	22..31
FT		/note= "epitope-bearing region"
FT	Peptide	35..44
FT		/note= "epitope-bearing region"
FT	Peptide	71..81
FT		/note= "epitope-bearing region"
FT	Peptide	91..107
FT		/note= "epitope-bearing region"
FT	Peptide	119..128
FT		/note= "epitope-bearing region"
FT	Peptide	138..147
FT		/note= "epitope-bearing region"
FT	Peptide	155..167
FT		/note= "epitope-bearing region"
FT	Peptide	193..203
FT		/note= "epitope-bearing region"
FT	Peptide	206..215
FT		/note= "epitope-bearing region"
FT	Peptide	227..237
FT		/note= "epitope-bearing region"
FT	Peptide	243..252

```
FT                   /note= "epitope-bearing region"
PN   WO9854199-A1.
PD   03-DEC-1998.
PF   27-MAY-1998; U10728.
PR   28-MAY-1997; US-048000.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Ebner R, Moore PA, Ruben SM;
DR   WPI; 99-070207/06.
DR   N-PSDB; V99636.
PT   New tissue plasminogen activator-like protease - useful in the
PT   diagnosis and treatment of circulatory system-related disorders
PS   Claim 1; Page 56-57; 76pp; English.
CC   This is the amino acid sequence of tissue plasminogen activator-like
CC   protease (t-PALP), a novel member of the serine protease family
CC   that shares sequence homology to human tissue plasminogen activator
CC   (see W87770).  The t-PALP sequence was deduced from a cDNA clone
CC   (see V99636) derived from activated monocytes.  The 2.5 kb t-PALP
CC   message has also been detected in heart, brain, lung, placenta,
CC   liver, skeletal muscle, kidney, pancreas, spleen, thymus, prostate,
CC   testis, ovary, small intestine, colon and peripheral blood
CC   leukocytes.  Isolated nucleic acids encoding amino acids -21 to
CC   242, -20 to 242, 1-242, 4-63 (kringle domain) and 64-242 (protease
CC   domain) of t-PALP, or encoding epitope-bearing portions of t-PALP,
CC   are also claimed, as are recombinant vectors, host cells, and
CC   methods for producing t-PALP polypeptides.  t-PALP may be used to
CC   detect and treat disorders related to the circulatory system, and
CC   to identify agonists and antagonists of t-PALP activity.  The
CC   homology between t-PALP and tPA indicates that t-PALP may be
CC   involved in the regulation of normal and abnormal clotting
CC   in e.g. stroke, deep-vein thrombosis, peripheral arterial
CC   occlusion, pulmonary embolism and myocardiothrombosis.
SQ   Sequence   263 AA;

  Query Match          100.0%;  Score 1883;  DB 39;  Length 263;
  Best Local Similarity 100.0%;  Pred. No. 1.04e-174;
  Matches   263;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  -   1 mllawvqaflvsnmllaeaygsggcfwdnghlyredqtspapglrclnwldaqsglasap 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 MLLAWVQAFLVSNMLLAEAYGSGGCFWDNGHLYREDQTSPAPGLRCLNWLDAQSGLASAP 60

Db     61 vsgagnhsycrnpdedprgpwcyvsgeagvpekrpcedlrcpettsqalpaftteiqeas 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 VSGAGNHSYCRNPDEDPRGPWCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEAS 120

Db    121 egpgadevqvfapanalparseaaavqpvigisqrvrmnskekkdlgtlgyvlgitmmvi 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121 EGPGADEVQVFAPANALPARSEAAAVQPVIGISQRVRMNSKEKKDLGTLGYVLGITMMVI 180

Db    181 iiaigagiilgysykrgkdlkeqhdqkvceremqritlplsaftnptceivdektvvvht 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    181 IIAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRITLPLSAFTNPTCEIVDEKTVVVHT 240

Db    241 sqtpvdpqegstplmgqagtpga 263
          |||||||||||||||||||||||
Qy    241 SQTPVDPQEGSTPLMGQAGTPGA 263


RESULT    2
ID   W72641 standard; peptide; 39 AA.
AC   W72641;
DT   05-JAN-1999  (first entry)
DE   Nervous glia cell growth factor N-terminal peptide #2.
KW   Nervous glia cell growth factor; human; urine; secretion promoter;
KW   choline acetyltransferase activity enhancer; nervous disease.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 25
FT                   /note= "unspecified"
FT   Misc_difference 29
FT                   /note= "unspecified"
PN   J10265498-A.
PD   06-OCT-1998.
PF   24-MAR-1997; 090305.
PR   24-MAR-1997; JP-090305.
PA   (NICH-) JAPAN CHEM RES CO LTD.
DR   WPI; 98-589719/50.
PT   Nervous glia cell growth factor derived from human urine - used for
PT   treatment of nervous diseases
PS   Claim 3; Fig 7; 14pp; Japanese.
CC   The present invention describes nervous glia cell growth factor, which
CC   is purified from human urine by ultrafiltration, salting-out by ammonium
CC   sulphate, gel filtration, ion exchange chromatography and reversed phase
CC   chromatography, and has a M.W. of 29 kDa by sodium dodecyl sulphate-
CC   polyacrylamide gel electrophoresis. Also described are: (1) a secretion
CC   promoter for the nerve growth factor of glia cell consisting of the
CC   above growth factor, an enhancer for choline acetyltransferase activity
CC   of neuron consisting of the above growth factor; and (2) DNA encoding
CC   nervous glia growth factor containing a DNA sequence coding the amino
CC   acid sequence shown by the two 39 amino acid sequences as given in
CC   W72640 and W72641, which are identical, except one starts with Tyr and
CC   the other with Ser (i.e. they are from different DNA transcripts). The
CC   glia cell growth factor can be prepared in a large amount and the factor
CC   can be used for the treatment of nervous diseases.
SQ   Sequence   39 AA;

  Query Match          13.6%;  Score 256;  DB 36;  Length 39;
  Best Local Similarity 92.3%;  Pred. No. 3.64e-13;
  Matches   36;  Conservative    0;  Mismatches   3;  Indels   0;  Gaps   0;

Db      1 sgggfwdnghlyredqtspapglrxlnwxdaqsglasap 39
          ||| ||||||||||||||||||| ||| ||||||||||
Qy     22 SGGCFWDNGHLYREDQTSPAPGLRCLNWLDAQSGLASAP 60


RESULT    3
ID   W72640 standard; peptide; 39 AA.
AC   W72640;
DT   05-JAN-1999  (first entry)
DE   Nervous glia cell growth factor N-terminal peptide #1.
KW   Nervous glia cell growth factor; human; urine; secretion promoter;
KW   choline acetyltransferase activity enhancer; nervous disease.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 25
FT                   /note= "unspecified"
FT   Misc_difference 29
FT                   /note= "unspecified"
PN   J10265498-A.
PD   06-OCT-1998.
PF   24-MAR-1997; 090305.
PR   24-MAR-1997; JP-090305.
PA   (NICH-) JAPAN CHEM RES CO LTD.
DR   WPI; 98-589719/50.
PT   Nervous glia cell growth factor derived from human urine - used for
PT   treatment of nervous diseases
PS   Claim 2; Fig 6; 14pp; Japanese.
CC   The present invention describes nervous glia cell growth factor, which
CC   is purified from human urine by ultrafiltration, salting-out by ammonium
CC   sulphate, gel filtration, ion exchange chromatography and reversed phase
CC   chromatography, and has a M.W. of 29 kDa by sodium dodecyl sulphate-
CC   polyacrylamide gel electrophoresis. Also described are: (1) a secretion
CC   promoter for the nerve growth factor of glia cell consisting of the
CC   above growth factor, an enhancer for choline acetyltransferase activity
CC   of neuron consisting of the above growth factor; and (2) DNA encoding
CC   nervous glia growth factor containing a DNA sequence coding the amino
CC   acid sequence shown by the two 39 amino acid sequences as given in
CC   W72640 and W72641, which are identical, except one starts with Tyr and
CC   the other with Ser (i.e. they are from different DNA transcripts). The
CC   glia cell growth factor can be prepared in a large amount and the factor
CC   can be used for the treatment of nervous diseases.
SQ   Sequence   39 AA;

  Query Match          13.5%;  Score 255;  DB 36;  Length 39;
  Best Local Similarity 90.2%;  Pred. No. 4.50e-13;
  Matches   37;  Conservative    0;  Mismatches   2;  Indels   2;  Gaps   2;
```

```
Db      1 yg-gg-fwdnghlyredqtspapglrxlnwxdaqsglasap 39
          || || ||||||||||||||||||| ||| ||||||||||
Qy     20 YGSGGCFWDNGHLYREDQTSPAPGLRCLNWLDAQSGLASAP 60
```

RESULT 4
ID R83959 standard; protein; 812 AA.
AC R83959;
DT 10-MAR-1996 (first entry)
DE Complete mouse plasminogen molecule.
KW Angiostatin; plasminogen; endothelial inhibitor; therapeutic;
KW gene therapy.
OS Mus musculus.
FH Key Location/Qualifiers
FT protein 1..812
FT /note= "plasminogen"
FT protein 98..436
FT /note= "angiostatin"
PN WO9529242-A1.
PD 02-NOV-1995.
PF 26-APR-1995; U05107.
PR 26-APR-1994; US-248629.
PR 20-OCT-1994; US-326785.
PA (CHIL-) CHILDRENS MEDICAL CENT.
PI Cao Y, Folkman MJ, O'Reilly MS, Sim KL;
DR WPI; 95-382990/49.
PT Endothelial inhibitor Angiostatin - useful to treat angiogenic
PT mediated disease esp. angiogenesis and cancer.
PS Disclosure; Fig 1; 108pp; English.
CC Angiostatin (see R83960) is a plasminogen fragment starting at
CC amino acid 98 of the complete plasminogen molecule. Preferably,
CC angiostatin has an amino acid sequence similar to that of the
CC plasminogen fragment. Angiostatin is an endothelial inhibitor,
CC which reversibly inhibits proliferation of endothelial cells and
CC thereby inhibits angiogenesis. It is useful in the treatment of
CC a human or animal with angiogenic mediated disease e.g. arthritis,
CC macular degeneration, diabetic retinopathy or cancer. Cells
CC comprising angiostatin-coding sequences are useful for gene therapy
CC of primary tumors.
SQ Sequence 812 AA;

Query Match 9.8%; Score 184; DB 15; Length 812;
Best Local Similarity 27.8%; Pred. No. 1.03e-06;
Matches 35; Conservative 38; Mismatches 43; Indels 10; Gaps 9;

```
Db    377 cyqsdgqsyrgtssttitgkkcqswaamfphrhsktpenfpdaglemnycrnpdgd-kgp 435
          |:  :|: ||   ::   | :| :|    ::  : ::   : || : :|||||| | :||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLD--AQS-GLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db    436 wcytt-dpsvr-weycnlkrcsetgg-svvelptvsqeps-gpsdsetdcmygngkdyrg 491
          ||| : :::|    |:  ||:|| : ::  : |  ||:| ||:  | : ::: ::   :
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQ-VFAPANALPA 139

Db    492 ktavta 497
          ::  :|
Qy    140 RSEAAA 145
```

RESULT 5
ID W07585 standard; protein; 812 AA.
AC W07585;
DT 24-JUN-1997 (first entry)
DE Murine plasminogen sequence.
KW angiostatin; plasminogen; kringle; angiogenesis; cancer; arthritis;
KW macular degeneration; diabetic retinopathy.
OS Mus musculus.
PN WO9635774-A2.
PD 14-NOV-1996.
PF 26-APR-1996; U05856.
PR 26-APR-1995; US-429743.
PR 22-FEB-1996; US-605598.
PR 08-MAR-1996; US-612788.
PA (CHIL-) CHILDRENS MEDICAL CENT.
PI Cao Y, Folkman MJ, Lin J, Oreilly MS, Sim KL;
DR WPI; 96-518662/51.
PT Use of angiostatin fragments or aggregates - for inhibiting
PT endothelial cell proliferation and treating angiogenesis-mediated
PT diseases, e.g. cancer, arthritis or diabetic retinopathy
PS Disclosure; Fig 1; 203pp; English.
CC The invention relates to new methods and compositions for
CC inhibiting endothelial cell proliferation, using as active component
CC an angiostatin fragment, a combination of angiostatin fragments, or
CC aggregate angiostatin. The fragment is preferably derived from murine,
CC human, Rhesus, porcine or bovine plasminogen and is a kringle 1,
CC kringle 2, kringle 3, kringle 2-3, kringle 1-3, kringle 1-2, kringle
CC 1-4 or kringle 1-4BKLS protein. The aggregate angiostatin has a Mol. Wt.
CC of 45-65 kD and is derived from a plasminogen fragment beginning at
CC approximately amino acid number 98 of murine, human, Rhesus, porcine or
CC bovine plasminogen. The active component can be used for treating
CC angiogenesis-mediated diseases such as cancer, arthritis, macular
CC degeneration and diabetic retinopathy. It can also be used to develop
CC antibodies for use in diagnosis, detection and therapy.
CC The present sequence is the whole murine plasminogen sequence from
CC which fragments or aggregates may be derived for use in the invention.
SQ Sequence 812 AA;

Query Match 9.8%; Score 184; DB 21; Length 812;
Best Local Similarity 27.8%; Pred. No. 1.03e-06;
Matches 35; Conservative 38; Mismatches 43; Indels 10; Gaps 9;

```
Db    377 cyqsdgqsyrgtssttitgkkcqswaamfphrhsktpenfpdaglemnycrnpdgd-kgp 435
          |:  :|: ||   ::   | :| :|    ::  : ::   : || : :|||||| | :||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLD--AQS-GLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db    436 wcytt-dpsvr-weycnlkrcsetgg-svvelptvsqeps-gpsdsetdcmygngkdyrg 491
          ||| : :::|    |:  ||:|| : ::  : |  ||:| ||:  | : ::: ::   :
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQ-VFAPANALPA 139

Db    492 ktavta 497
          ::  :|
Qy    140 RSEAAA 145
```

RESULT 6
ID R89197 standard; protein; 655 AA.
AC R89197;
DT 08-AUG-1996 (first entry)
DE Human hepatocellular growth factor single chain precursor protein.
KW Mature protein; inactive; single chain; hepatocellular growth factor;
KW liver; human; precursor; dimerisation; double chain; serine protease;
KW hepatitis.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 356..655
FT /note= "mature peptide"
PN J08027026-A.
PD 30-JAN-1996.
PF 22-JUL-1994; 171207.
PR 22-JUL-1994; JP-171207.
PA (MITU) MITSUBISHI CHEM CORP.
DR WPI; 96-136206/14.
PT New preventative and therapeutic cpds contg. a 34 kD serine protease
PT - useful for treatment of liver diseases e.g. hepatitis.
PS Claim 4; Page 6-8; 8pp; Japanese.
CC This is the amino acid sequence of the precursor protein of the inactive
CC single chain form of a hepatocellular growth factor. The mature protein
CC (R89196) has a mol. wt. of around 34 kD and is derived from the 96 kD
CC precursor protein. The mature protein corresp. to residues 356-655 of
CC the precursor protein. The inactive single chain form of the growth
CC factor is activated by dimerisation of the mature protein. The active
CC protein is a serine protease which can be used for the treatment of
CC hepatitis.
SQ Sequence 655 AA;

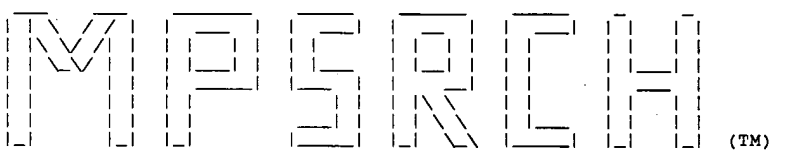
Query Match 9.6%; Score 181; DB 17; Length 655;

```
  Best Local Similarity  40.5%;  Pred. No. 1.88e-06;
  Matches    30;  Conservative   19;  Mismatches  17;  Indels   8;  Gaps   4;

Db    286 cflgngtgyrgvastsasglsclawnsdllyqelhvdsvgaaallglgphaycrnpdnde 345
          ||  ||  ||   :::|:|| || |    |  : :  ::::|:: | | |:||||||:|
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW----L--DAQ-SGLASAPVSGAGNHSYCRNPDEDP 77

Db    346 r-pwcyvvkdsals 358
          | |||||  :::::
Qy     78 RGPWCYVSGEAGVP 91


RESULT    7
ID   R53962 standard; Protein; 655 AA.
AC   R53962;
DT   06-JAN-1995  (first entry)
DE   Hepatocyte growth factor converting protease.
KW   Hepatocyte growth factor; protease; cleavage; active; inactive;
KW   precursor.
OS   Homo sapiens.
PN   EP-596524-A.
PD   11-MAY-1994.
PF   05-NOV-1993; 117988.
PR   05-NOV-1992; JP-296133.
PR   20-NOV-1992; JP-312234.
PR   20-NOV-1992; JP-312242.
PA   (SHIM/) SHIMOMURA T.
PA   (MITU ) MITSUBISHI KASEI CORP.
PI   Kitamura N,  Miyazawa K,  Morimoto Y,  Shimomura T;
PI   Yamada K;
DR   WPI; 94-152921/19.
DR   N-PSDB; Q63951.
PT   Hepatocyte growth factor converting protease and precursor and
PT   gene encoding them - for producing active two chain HGF from
PT   inactive single chain HGF
PS   Claim 12; Page 21-24; 30pp; English.
CC   Hepatocyte growth factor converting protease is capable of converting
CC   inactive single chain hepatocyte growth factor (HGF) into active two
CC   chain HGF by cleavage at a specific site.
SQ   Sequence   655 AA;

  Query Match             9.6%;  Score 181;  DB 10;  Length 655;
  Best Local Similarity  40.5%;  Pred. No. 1.88e-06;
  Matches    30;  Conservative   19;  Mismatches  17;  Indels   8;  Gaps   4;

Db    286 cflgngtgyrgvastsasglsclawnsdllyqelhvdsvgaaallglgphaycrnpdnde 345
          ||  ||  ||   :::|:|| || |    |  : :  ::::|:: | | |:||||||:|
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW----L--DAQ-SGLASAPVSGAGNHSYCRNPDEDP 77

Db    346 r-pwcyvvkdsals 358
          | |||||  :::::
Qy     78 RGPWCYVSGEAGVP 91


RESULT    8
ID   R66600 standard; Protein; 701 AA.
AC   R66600;
DT   13-FEB-1995  (first entry)
DE   Mouse L5/3 tumour suppressor protein.
KW   Mouse L5/3 gene; small lung cell carcinoma; tumour suppression;
KW   chromosome 3; 3p21; D3F15S2 locus; hepatocyte growth factor;
KW   renal cell carcinoma; Von Hipple-Lindau syndrome; predisposition;
KW   kringle domain.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   peptide         1..16
FT                   /label= signal_peptide_(16-31)
FT                   /note= "putative"
FT   misc_difference 4
FT                   /label= polymorphic_site
FT                   /note= "Pro corresponds to CCG codon in cDNA;
FT                   in the genomic DNA, codon 19 is CAG (Gln)"
FT   protein         17..701
FT                   /note= "putative protein contains 4 kringle domains
FT                   followed by a serine protease-like domain"
FT   modified_site   57..59
FT                   /label= N-glycosylation_site
FT                   /note= "potential"
FT   modified_site   158..160
FT                   /label= N-glycosylation_site
FT                   /note= "potential"
FT   modified_site   290..292
FT                   /label= N-glycosylation_site
FT                   /note= "potential"
FT   modified_site   605..607
FT                   /label= N-glycosylation_site
FT                   /note= "potential"
PN   US5315000-A.
PD   24-MAY-1994.
PF   14-MAY-1992; 882925.
PR   14-MAY-1992; US-882925.
PA   (CHIL-) CHILDRENS HOSPITAL MEDICAL CENT.
PI   Degen SJ;
DR   WPI; 94-166645/20.
DR   N-PSDB; Q79726.
PT   DNA from D3F15S2 locus of human chromosome 3 - encoding novel
PT   growth factor, L5/3, useful as probe for detecting
PT   pre-deposition towards cancer
PS   Disclosure; Columns 27-32; 31pp; English.
CC   A mouse liver cDNA library was screened with cDNA coding for human
CC   L5/3; the L5/3 gene, located at the D3F15S2 locus of human
CC   chromosome 3, codes for a protein composed of 4 kringle domains,
CC   followed by a serine protease-like domain. The longest murine clone
CC   to be isolated (pML5-2, Q79726) was not full-length. The open
CC   reading frame was present at the 5'end of the sequence with no
CC   codon for the initiator methionine in-frame with the coding
CC   sequence. After determination of the sequence of the mouse gene it
CC   was determined that the cDNA lacked 44bp of coding and 94bp of
CC   non-coding sequence at its 5'-end.
SQ   Sequence   701 AA;

  Query Match             9.4%;  Score 177;  DB 10;  Length 701;
  Best Local Similarity  41.3%;  Pred. No. 4.17e-06;
  Matches    26;  Conservative    9;  Mismatches  26;  Indels   2;  Gaps   2;

Db     95 cimdngvsyrgtvartagglpcqawsrrfpndhkytptpkngleenfcrnpdgdprgpwc 154
          |: |||  ||   :  | || |  |     :    :| :  | : ::|||||| |||||||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLDA-QSGLASAPVSGAG-NHSYCRNPDEDPRGPWC 82

Db    155 ytt 157
          | :
Qy     83 YVS 85


RESULT    9
ID   W14271 standard; Protein; 701 AA.
AC   W14271;
DT   21-JUL-1997  (first entry)
DE   Mouse growth factor L5/3 partial cDNA clone ML5-2 encoded protein.
KW   Mouse; growth factor; foetal; liver; probe; bovine; prothrombin; locus;
KW   polymorphism; transition; exon; intron; chromosome; kringle domain;
KW   cell growth; tumour suppressor; hepatocyte growth factor; regeneration.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   peptide         1..16
FT                   /note= "partial signal peptide sequence"
FT   misc_difference 19
FT                   /note= "amino acid residue is Gln in this position
FT                   in the protein encoded by the genomic sequence
FT                   (T62442); this may be due to a polymorphism"
FT   protein         17..701
FT                   /note= "mature protein"
PN   US5606029-A.
PD   25-FEB-1997.
PF   14-MAY-1992; 882925.
PR   14-MAY-1992; US-882925.
```

```
PR   18-JAN-1994; US-184012.
PA   (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
PI   Degen SJ;
DR   WPI; 97-153621/14.
DR   N-PSDB; T62441.
PT   Human growth factor protein L5/3 - useful for altering cell growth,
PT   e.g. as tumour suppressor
PS   Disclosure; Column 27-32; 34pp; English.
CC   This is the amino acid sequence encoded by the insert isolated from
CC   clone ML5-2 and constitutes part of a mouse growth factor designated L5/3
CC   (W14272).  The encoding sequence was isolated from a lambda-gt10 mouse
CC   liver cDNA library using a fragment of the corresponding human cDNA
CC   sequence (T62436) as a probe.  The fragment presented here is a partial
CC   sequence which lacks 44 bp of the coding region and 94 bp of the
CC   non-coding region.  A 1450 bp fragment of this sequence (nucleotides
CC   738-2188) covering 8 amino acids of the second kringle domain and all of
CC   the third and fourth kringle domains and the serine protease-like domain,
CC   was used to screen a mouse liver genomic DNA library to isolate the full
CC   length genomic sequence (T62442).  The protein can be used to alter cell
CC   growth (as a growth factor or tumour suppressor) and has similar
CC   properties to the hepatocyte growth factor that is actively involved in
CC   liver regeneration.
SQ   Sequence   701 AA;

  Query Match             9.4%;  Score 177;  DB 22;  Length 701;
  Best Local Similarity  41.3%;  Pred. No. 4.17e-06;
  Matches    26;  Conservative    9;  Mismatches  26;  Indels   2;  Gaps   2;

Db     95 cimdngvsyrgtvartagglpcqawsrrfpndhkytptpkngleenfcrnpdgdprgpwc 154
          |: |||  ||   :  | || |  |     :    :| :  | : ::|||||| |||||||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLDA-QSGLASAPVSGAG-NHSYCRNPDEDPRGPWC 82

Db    155 ytt 157
          | :
Qy     83 YVS 85


RESULT   10
ID   R66601 standard; Protein; 716 AA.
AC   R66601;
DT   14-FEB-1995  (first entry)
DE   Mouse L5/3 tumour suppressor protein (from genomic sequence).
KW   Mouse L5/3 gene; small lung cell carcinoma; tumour suppression;
KW   chromosome 3; 3p21; D3F15S2 locus; hepatocyte growth factor;
KW   renal cell carcinoma; Von Hipple-Lindau syndrome; predisposition;
KW   kringle domain.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   peptide         1..31
FT                   /label= signal_peptide
FT                   /note= "putative"
FT   misc_difference 19
FT                   /label= polymorphic_site
FT                   /note= "Gln corresponds to CAG codon in genomic DNA;
FT                   in the cDNA, codon 19 is CCG (Pro)(R66600)"
FT   protein         32..716
FT                   /note= "putative protein contains 4 kringle domains
FT                   followed by a serine protease-like domain"
FT   modified_site   72..74
FT                   /label= N-glycosylation_site
FT                   /note= "potential"
FT   modified_site   173..175
FT                   /label= N-glycosylation_site
FT                   /note= "potential"
FT   modified_site   305..307
FT                   /label= N-glycosylation_site
FT                   /note= "potential"
FT   modified_site   620..622
FT                   /label= N-glycosylation_site
FT                   /note= "potential"
PN   US5315000-A.
PD   24-MAY-1994.
PF   14-MAY-1992; 882925.
PR   14-MAY-1992; US-882925.
PA   (CHIL-) CHILDRENS HOSPITAL MEDICAL CENT.
PI   Degen SJ;
DR   WPI; 94-166645/20.
DR   N-PSDB; Q79727.
PT   DNA from D3F15S2 locus of human chromosome 3 - encoding novel
PT   growth factor, L5/3, useful as probe for detecting
PT   pre-deposition towards cancer
PS   Disclosure; Columns 33-42; 31pp; English.
CC   The sequence of mouse genomic DNA coding for the L5/3 tumour
CC   suppressor protein is composed of 18 exons separated by 17
CC   intervening sequences. There is only one difference found between
CC   the cDNA (Q79726) and genomic DNA (Q79727) coding sequences which
CC   results in the substitution of a Gln in the gene to a Pro in the
CC   cDNA at amino acid position 19. The putative mouse protein has the
CC   same domain structure as its human homologue with four kringle
CC   domains followed by a serine protease-like domain.
SQ   Sequence   716 AA;

  Query Match             9.4%;  Score 177;  DB 10;  Length 716;
  Best Local Similarity  41.3%;  Pred. No. 4.17e-06;
  Matches    26;  Conservative    9;  Mismatches  26;  Indels   2;  Gaps   2;

Db    110 cimdngvsyrgtvartagglpcqawsrrfpndhkytptpkngleenfcrnpdgdprgpwc 169
          |: |||  ||   :  | || |  |     :    :| :  | : ::|||||| |||||||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLDA-QSGLASAPVSGAG-NHSYCRNPDEDPRGPWC 82

Db    170 ytt 172
          | :
Qy     83 YVS 85


RESULT   11
ID   W14272 standard; Protein; 716 AA.
AC   W14272;
DT   21-JUL-1997  (first entry)
DE   Mouse growth factor L5/3 complete protein.
KW   Mouse; growth factor; foetal; liver; probe; bovine; prothrombin; locus;
KW   polymorphism; transition; exon; intron; chromosome; kringle domain;
KW   cell growth; tumour suppressor; hepatocyte growth factor; regeneration.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   peptide         1..31
FT                   /note= "signal peptide"
FT   misc_difference 19
FT                   /note= "amino acid residue is Pro at this position
FT                   in the protein encoded by the cDNA clone ML5-2
FT                   (T62441); this may be due to a polymorphism"
FT   protein         32..716
FT                   /note= "mature protein"
FT   modified_site   72
FT                   /note= "N-linked glycosylation site"
FT   modified_site   173
FT                   /note= "N-linked glycosylation site"
FT   modified_site   305
FT                   /note= "N-linked glycosylation site"
FT   modified_site   624
FT                   /note= "N-linked glycosylation site"
PN   US5606029-A.
PD   25-FEB-1997.
PF   14-MAY-1992; 882925.
PR   14-MAY-1992; US-882925.
PR   18-JAN-1994; US-184012.
PA   (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
PI   Degen SJ;
DR   WPI; 97-153621/14.
DR   N-PSDB; T62442.
PT   Human growth factor protein L5/3 - useful for altering cell growth,
PT   e.g. as tumour suppressor
PS   Disclosure; Column 33-42; 34pp; English.
CC   This is the amino acid sequence of the wild type mouse growth factor
CC   designated L5/3. The protein sequence differs from that encoded by the
CC   cDNA clone (T62441) at position 19; in this sequence a Gln, in the cDNA
```

```
CC   clone a Pro. This difference may be due to a polymorphism at this codon.
CC   The full length mouse gene contains 18 exons and encodes a protein having
CC   a molecular weight 80 kD. The protein can be used to alter cell growth
CC   (as a growth factor or tumour suppressor) and has similar properties to
CC   the hepatocyte growth factor that is actively involved in liver
CC   regeneration.
SQ   Sequence   716 AA;

  Query Match            9.4%;  Score 177;  DB 22;  Length 716;
  Best Local Similarity  41.3%;  Pred. No. 4.17e-06;
  Matches    26;  Conservative    9;  Mismatches  26;  Indels   2;  Gaps   2;

Db    110 cimdngvsyrgtvartagglpcqawsrrfpndhkytptpkngleenfcrnpdgdprgpwc 169
          |: ||| ||   :  | || |  |     :    :| :  | : ::||||| |||||||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLDA-QSGLASAPVSGAG-NHSYCRNPDEDPRGPWC 82

Db    170 ytt 172
          | :
Qy     83 YVS 85


RESULT   12
ID   W46423 standard; protein; 217 AA.
AC   W46423;
DT   13-MAY-1998  (first entry)
DE   Bovine macrophage stimulating protein (MSP).
KW   p55; p22; p75; human; macrophage stimulating protein; MSP;
KW   bovine analogue; colony formation; crypt cell; proliferation;
KW   crypt stem cell; intestinal epithelial cell; treatment; disorder;
KW   gastrointestinal tract lining.
OS   Bos sp.
PN   WO9735606-A1.
PD   02-OCT-1997.
PF   25-MAR-1997; U05216.
PR   26-MAR-1996; US-622720.
PA   (AMGE-) AMGEN INC.
PI   Zhang K;
DR   WPI; 97-489386/45.
PT   Treatment of disorders of gastrointestinal tract lining in mammals -
PT   which result from chemotherapy, radiation therapy, inflammatory
PT   bowel disease, ulcers or infection
PS   Disclosure; Fig 8; 57pp; English.
CC   The present sequence represents the bovine analogue of macrophage
CC   stimulating protein (MSP). MSP stimulates colony formation by crypt
CC   cells and may also stimulate proliferation of crypt stem cells. In this
CC   respect, MSP is useful for elevating production of all intestinal
CC   epithelial cell types. Disorders of the gastrointestinal tract lining in
CC   mammals can be treated by administering MSP. The method can be used to
CC   treat disorders of the gastrointestinal tract tract lining in a mammal
CC   which result from chemotherapy, radiation therapy, inflammatory bowel
CC   disease, ulcers, infection, ulcerative colitis or Crohn's disease.
SQ   Sequence   217 AA;

  Query Match            9.3%;  Score 176;  DB 28;  Length 217;
  Best Local Similarity  32.4%;  Pred. No. 5.09e-06;
  Matches    23;  Conservative   20;  Mismatches  23;  Indels   5;  Gaps   5;

Db     87 cyhgagelyrgsvsktrkgircqnwsaetphkpqfkhtsaphtpleenfcrnpdgdshgp 146
          |:   | |||   :    |:|| ||  :: : : :  :: : :  : ::|||||  |::||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW-LDA-Q-SGLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db    147 wcytt-dpgtp 156
          ||| : ::| |
Qy     81 WCYVSGEAGVP 91


RESULT   13
ID   R05433 standard; protein; 380 AA.
AC   R05433;
DT   30-JUL-1990  (first entry)
DE   cPA-P2 Hybrid plasminogen activator.
KW   Plasminogen activator; fibrin; urokinase; thromboembolic
KW   disease; ds.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   domain          1..23
FT                   /label=Secretory leader.
FT   domain          24..102
FT                   /label=Plasminogen Kringle 1
FT   domain          103..116
FT                   /label=Urokinase linker
FT   domain          117..380
FT                   /label=ScuPA protease domain.
PN   WO9001332-A.
PD   22-FEB-1990.
PF   10-AUG-1988; 02771.
PR   10-AUG-1988; WO-U02771.
PA   (CETU) Cetus Corp.
PI   Halluin AP;
DR   WPI; 90-083374/11.
DR   N-PSDB; Q02301.
PT   Compsn. contg. plasminogen activator conjugated to heparin component -
PT   used for treatment of thromboembolic disease, with longer half
PT   life and improved targeting.
PS   Disclosure; p; English.
CC   Gene encodes hybrid plasminogen activator (PA) comprising Kringle 1, an
CC   urokinase linker, and an urokinase protease domain wherin glycine residue
CC   at position 158 is replaced with a lysine.
CC   The compound is used to treat thromboembolic disease esp. with myocardial
CC   infarction, has a longer half-life than free PA and targets the heparin
CC   site of thrombus or embolism reducing the risk of reocclusion.
SQ   Sequence   380 AA;

  Query Match            9.3%;  Score 175;  DB 1;  Length 380;
  Best Local Similarity  34.8%;  Pred. No. 6.21e-06;
  Matches    39;  Conservative   23;  Mismatches  39;  Indels  11;  Gaps   8;

Db      1 malaitlalllllllllpgcwase-cktgdgknyrgtmsktkngitcqkwsstsphrprfs 59
            | ||   |:|:  :||:   :|  |   :|  ||   :    |: | :|    ::   ::
Qy      1 MLLAWVQAFLVSNMLLAEAYGSGGCFWDNGHLYREDQTSPAPGLRCLNW--LDAQSG-LA 57

Db     60 pathpsegleenycrnpdndpqgpwcytt-d---pekrydycdileceeeps 107
          :|  :: | : :|||||:||:|||| : :   ||||   |: | | |  |
Qy     58 SAPVSGAG-NHSYCRNPDEDPRGPWCYVSGEAGVPEKR-P-CEDLRCPETTS 106


RESULT   14
ID   R32710 standard; Protein; 504 AA.
AC   R32710;
DT   16-JUN-1993  (first entry)
DE   Haematopoietic stem cell multiplier.
KW   Bone marrow deficiencies; cancer therapy;  tumour; carcinoma;
KW   bone marrow transplants.
PN   WO9303061-A.
PD   18-FEB-1993.
PF   24-JUL-1992; J00949.
PR   26-JUL-1991; JP-187470.
PR   26-JUL-1991; JP-187481.
PA   (TORA ) TORAY IND INC.
PI   Kawano G,  Kojima K,  Komiyama A,  Kubo T,  Nakahata T;
PI   Sano E,  Sudot,  Tanaka R;
DR   WPI; 93-076441/09.
DR   N-PSDB; Q37308.
PT   Haematopoietic stem cell multiplier comprising IL-3 and IL-7 -
PT   used for treatment and prevention of bone marrow disorders e.g.
PT   after cancer therapy or bone marrow transplants
PS   Disclosure; Page 60; 90pp; Japanese.
CC   This sequence is haematopoietic stem cell multiplier of mol.
CC   wt. 60,000, and having N-terminal sequence R32709. It can be used
CC   in the treatment (claimed) of bone marrow deficiencies eg. after cancer
CC   therapy or bone marrow transplants.
SQ   Sequence   504 AA;

  Query Match            9.1%;  Score 171;  DB 6;  Length 504;
  Best Local Similarity  31.5%;  Pred. No. 1.38e-05;
  Matches    29;  Conservative   22;  Mismatches  35;  Indels   6;  Gaps   4;
```

```
Db    386 cyrgngknymgnlsqtrsgltcsmwdknmedlhrhifwepdasklnenycrnpdddahgp 445
          |:: ||  |  : :   :|| |  |   ::  :  :   | ::  | :||||||:|::||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW---LDA-QSGLASAPVSGAGNHSYCRNPDEDPRGP 80

Db    446 wcy-tgnplip-wdycpisrcegdttptnsqf 475
          ||| :|:: :|    |   ||   |: : : |
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAF 112


RESULT   15
ID   R21976 standard; Protein; 728 AA.
AC   R21976;
DT   03-JUL-1992  (first entry)
DE   Human Hepatocyte growth factor.
KW   rhHGF; beta chain; rat HGF; HBC25; HAC19; ss.
OS   Homo sapiens.
PN   J04030000-A.
PD   31-JAN-1992.
PF   24-MAY-1990; 212818.
PR   05-JUN-1989; JP-142697.
PR   01-JAN-1990; JP-212818.
PR   24-MAY-1990; JP-134487.
PA   (TOYM ) TOYOBO KK.
DR   WPI; 92-085905/11.
DR   N-PSDB; Q22146.
PT   Recombinant human hepatocyte growth factor - for treatment and
PT   diagnosis of liver diseases
PS   Claim 2; Fig 4; 21pp; Japanese.
CC   A first cDNA library (I) was prepared from human liver RNA. The
CC   library was screened with the rat HGF beta chain coding sequence
CC   RBC1 (see Q22142). Clone HBC25 was isolated and sequenced (see
CC   Q22143). The partial human HGF coding sequence was itself used
CC   as a probe to screen a second cDNA library (II). A clone which
CC   partially overlapped with HBC25 was identified and designated
CC   HAC19 (see Q22144). The complete human HGF coding sequence could
CC   then be derived from the two overlapping sequences and the amino
CC   acid sequence of HGF deduced from it.
CC   See also Q22141 and Q22145.
SQ   Sequence   728 AA;

  Query Match            9.0%;  Score 170;  DB 4;  Length 728;
  Best Local Similarity  31.5%;  Pred. No. 1.68e-05;
  Matches    28;  Conservative    22;  Mismatches  33;  Indels   6;  Gaps   4;

Db    391 cyrgngknymgnlsqtrsgltcsmwnknmedlhrhifwepdasnlnenycrnpdddahgp 450
          |:: ||  |  : :   :|| |  |   ::  :  :   | ::  | :||||||:|::||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW---LDA-QSGLASAPVSGAGNHSYCRNPDEDPRGP 80

Db    451 wcy-tgnplip-wdycpisrcegdttpti 477
          ||| :|:: :|    |   ||   |: ::
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQAL 109


Search completed: Fri Sep 17 18:25:31 1999
Job time : 105 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 17 18:29:09 1999; MasPar time 4.08 Seconds
654.758 Million cell updates/sec
Tabular output not generated.

Title: >US-09-084-491A-2
Description: (1-263) from US09084491A.pep
Perfect Score: 1883
Sequence: 1 MLLAWVQAFLVSNMLLAEAY.....PVDPQEGSTPLMGQAGTPGA 263

Scoring table: PAM 150
Gap 11

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 31.020; Variance 137.816; scale 0.225

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	184	9.8	812	3	PCT-US95-0	Sequence 1, Applicatio	8.64e-07
2	184	9.8	812	2	US-08-612-	Sequence 1, Applicatio	8.64e-07
3	184	9.8	812	2	US-08-605-	Sequence 1, Applicatio	8.64e-07
4	184	9.8	812	2	US-08-452-	Sequence 1, Applicatio	8.64e-07
5	184	9.8	812	2	US-08-429-	Sequence 1, Applicatio	8.64e-07
6	184	9.8	812	1	US-08-451-	Sequence 1, Applicatio	8.64e-07
7	184	9.8	812	2	US-08-326-	Sequence 1, Applicatio	8.64e-07
8	184	9.8	812	1	US-08-248-	Sequence 1, Applicatio	8.64e-07
9	181	9.6	655	1	US-08-148-	Sequence 12, Applicati	1.54e-06
10	181	9.6	655	1	US-08-448-	Sequence 12, Applicati	1.54e-06
11	176	9.3	217	2	US-08-622-	Sequence 24, Applicati	4.04e-06
12	167	8.9	723	1	US-08-290-	Sequence 3, Applicatio	2.26e-05
13	167	8.9	723	1	US-08-290-	Sequence 2, Applicatio	2.26e-05
14	167	8.9	723	1	US-08-290-	Sequence 1, Applicatio	2.26e-05
15	167	8.9	723	1	US-08-404-	Sequence 1, Applicatio	2.26e-05
16	167	8.9	723	1	US-07-838-	Sequence 1, Applicatio	2.26e-05
17	167	8.9	728	1	US-07-815-	Sequence 2, Applicatio	2.26e-05
18	167	8.9	728	1	US-08-087-	Sequence 22, Applicati	2.26e-05
19	165	8.8	579	1	US-08-295-	Sequence 4, Applicatio	3.30e-05
20	165	8.8	579	3	PCT-US92-1	Sequence 4, Applicatio	3.30e-05
21	165	8.8	615	1	US-07-998-	Sequence 3, Applicatio	3.30e-05
22	165	8.8	615	3	PCT-US92-1	Sequence 3, Applicatio	3.30e-05
23	165	8.8	615	1	US-08-462-	Sequence 3, Applicatio	3.30e-05
24	165	8.8	615	1	US-08-463-	Sequence 3, Applicatio	3.30e-05
25	166	8.8	711	1	US-08-184-	Sequence 8, Applicatio	2.73e-05
26	166	8.8	711	3	PCT-US95-1	Sequence 2, Applicatio	2.73e-05
27	166	8.8	711	1	US-08-334-	Sequence 2, Applicatio	2.73e-05
28	161	8.6	790	1	US-08-469-	Sequence 54, Applicati	7.04e-05
29	161	8.6	791	2	US-08-643-	Sequence 1, Applicatio	7.04e-05
30	161	8.6	810	4	5200340-8	Patent No. 5200340.	7.04e-05
31	161	8.6	810	1	US-07-854-	Sequence 2, Applicatio	7.04e-05
32	161	8.6	810	1	US-08-147-	Sequence 29, Applicati	7.04e-05
33	161	8.6	814	1	US-08-750-	Sequence 1, Applicatio	7.04e-05
34	158	8.4	101	2	US-08-643-	Sequence 11, Applicati	1.24e-04
35	159	8.4	378	2	US-08-612-	Sequence 41, Applicati	1.03e-04
36	157	8.3	160	2	US-08-612-	Sequence 35, Applicati	1.50e-04
37	157	8.3	339	3	PCT-US95-0	Sequence 3, Applicatio	1.50e-04
38	157	8.3	339	2	US-08-429-	Sequence 3, Applicatio	1.50e-04
39	157	8.3	339	2	US-08-326-	Sequence 3, Applicatio	1.50e-04
40	157	8.3	339	1	US-08-248-	Sequence 3, Applicatio	1.50e-04
41	157	8.3	339	2	US-08-612-	Sequence 3, Applicatio	1.50e-04
42	157	8.3	339	1	US-08-451-	Sequence 3, Applicatio	1.50e-04
43	157	8.3	339	2	US-08-452-	Sequence 3, Applicatio	1.50e-04
44	157	8.3	352	2	US-08-612-	Sequence 40, Applicati	1.50e-04
45	157	8.3	378	2	US-08-612-	Sequence 42, Applicati	1.50e-04

ALIGNMENTS

RESULT 1
ID PCT-US95-05107-1 STANDARD; PRT; 812 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 1, Application PC/TUS9505107
XX
CC Sequence 1, Application PC/TUS9505107
CC GENERAL INFORMATION:
CC APPLICANT: THE CHILDREN'S MEDICAL CENTER, CORPORATION
CC TITLE OF INVENTION: Angiostatin and Method of Use
CC NUMBER OF SEQUENCES: 6
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Jones & Askew
CC STREET: 191 Peachtree Street, 37th Floor
CC CITY: Atlanta
CC STATE: Georgia
CC COUNTRY: U.S.
CC ZIP: 30303-1769
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/05107
CC FILING DATE:
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/248,629
CC FILING DATE: 26-APR-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/326,785
CC FILING DATE: 20-OCT-1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Johnson, James D.
CC REGISTRATION NUMBER: 31,771
CC REFERENCE/DOCKET NUMBER: 05213-0122
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 404-818-3700
CC TELEFAX: 404-818-3799
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 812 amino acids
CC TYPE: amino acid

```
CC         STRANDEDNESS:
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
CC       ORIGINAL SOURCE:
CC         ORGANISM:  Murine
SQ   SEQUENCE   812 AA; 90846 MW; 3411937 CN;

  Query Match             9.8%;  Score 184;  DB 3;  Length 812;
  Best Local Similarity  27.8%;  Pred. No. 8.64e-07;
  Matches    35;  Conservative   38;  Mismatches  43;  Indels  10;  Gaps   9;

Db    377 CYQSDGQSYRGTSSTTITGKKCQSWAAMFPHRHSKTPENFPDAGLEMNYCRNPDGD-KGP 435
          |:  :|: ||   ::   | :| :|    ::  : ::  : || : :|||||| | :||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLD--AQS-GLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db    436 WCYTT-DPSVR-WEYCNLKRCSETGG-SVVELPTVSQEPS-GPSDSETDCMYGNGKDYRG 491
          ||| : :::|     |:  ||:|| : ::  : |  ||:| ||:  | : ::: ::   :
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQ-VFAPANALPA 139

Db    492 KTAVTA 497
          ::  :|
Qy    140 RSEAAA 145


RESULT    2
ID   US-08-612-788-1     STANDARD;     PRT;   812 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08612788
XX
CC    Sequence 1, Application US/08612788
CC    Patent No. 5837682
CC     GENERAL INFORMATION:
CC       APPLICANT:  Folkman, M. Judah
CC       APPLICANT:  O'Reilly, Micheal
CC       APPLICANT:  Cao, Yihai
CC       APPLICANT:  Sim, B. Kim Lee
CC       TITLE OF INVENTION:  Angiostatin Fragments and Method of Use
CC       NUMBER OF SEQUENCES:  45
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Jones & Askew
CC         STREET:  191 Peachtree Street, 37th Floor
CC         CITY:  Atlanta
CC         STATE:  Georgia
CC         COUNTRY:  U.S.
CC         ZIP:  30303-1769
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/612,788
CC         FILING DATE:
CC         CLASSIFICATION:  514
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Warren, William L.
CC         REGISTRATION NUMBER:  36,714
CC         REFERENCE/DOCKET NUMBER:  05213-0126
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  404-818-3700
CC         TELEFAX:  404-818-3799
CC     INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  812 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
CC       ANTI-SENSE:  NO
CC       FRAGMENT TYPE:  N-terminal
CC       ORIGINAL SOURCE:
CC         ORGANISM:  Murine
CC       IMMEDIATE SOURCE:
CC         CLONE:  Plasminogen
SQ   SEQUENCE   812 AA; 90846 MW; 3411937 CN;

  Query Match             9.8%;  Score 184;  DB 2;  Length 812;
  Best Local Similarity  27.8%;  Pred. No. 8.64e-07;
  Matches    35;  Conservative   38;  Mismatches  43;  Indels  10;  Gaps   9;

Db    377 CYQSDGQSYRGTSSTTITGKKCQSWAAMFPHRHSKTPENFPDAGLEMNYCRNPDGD-KGP 435
          |:  :|: ||   ::   | :| :|    ::  : ::   : || : :|||||| | :||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLD--AQS-GLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db    436 WCYTT-DPSVR-WEYCNLKRCSETGG-SVVELPTVSQEPS-GPSDSETDCMYGNGKDYRG 491
          ||| : :::|     |:  ||:|| : ::  : |  ||:| ||:  | : ::: ::   :
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQ-VFAPANALPA 139

Db    492 KTAVTA 497
          ::  :|
Qy    140 RSEAAA 145


RESULT    3
ID   US-08-605-598B-1     STANDARD;     PRT;   812 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08605598B
XX
CC    Sequence 1, Application US/08605598B
CC    Patent No. 5861372
CC     GENERAL INFORMATION:
CC       APPLICANT:  Folkman, M. Judah
CC       APPLICANT:  Lin, Jie
CC       APPLICANT:  O'Reilly, Michael S.
CC       TITLE OF INVENTION:  Aggregate Angiostatin and Method of Use
CC       NUMBER OF SEQUENCES:  6
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Jones & Askew
CC         STREET:  191 Peachtree Street, 37th Floor
CC         CITY:  Atlanta
CC         STATE:  Georgia
CC         COUNTRY:  U.S.
CC         ZIP:  30303-1769
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/605,598B
CC         FILING DATE:  22-FEB-1996
CC         CLASSIFICATION:  514
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Warren, William L.
CC         REGISTRATION NUMBER:  36,714
CC         REFERENCE/DOCKET NUMBER:  05213-0127
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  404-818-3700
CC         TELEFAX:  404-818-3799
CC     INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  812 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
```

```
CC       ORIGINAL SOURCE:
CC         ORGANISM:  Murine Plasminogen
SQ   SEQUENCE   812 AA; 90846 MW; 3411937 CN;

  Query Match             9.8%;  Score 184;  DB 2;  Length 812;
  Best Local Similarity  27.8%;  Pred. No. 8.64e-07;
  Matches    35;  Conservative   38;  Mismatches  43;  Indels  10;  Gaps   9;

Db     377 CYQSDGQSYRGTSSTTITGKKCQSWAAMFPHRHSKTPENFPDAGLEMNYCRNPDGD-KGP 435
           |:  :|: ||   ::   | :| :|    ::  : ::   : || : :|||||| | :||
Qy      25 CFWDNGHLYREDQTSPAPGLRCLNWLD--AQS-GLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db     436 WCYTT-DPSVR-WEYCNLKRCSETGG-SVVELPTVSQEPS-GPSDSETDCMYGNGKDYRG 491
           ||| : :::|     |:  ||:|| : ::  : |  ||:| ||:  | : ::: ::   :
Qy      81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQ-VFAPANALPA 139

Db     492 KTAVTA 497
           ::  :|
Qy     140 RSEAAA 145


RESULT    4
ID   US-08-452-260-1     STANDARD;     PRT;   812 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08452260
XX
CC    Sequence 1, Application US/08452260
CC    Patent No. 5776704
CC     GENERAL INFORMATION:
CC       APPLICANT:  Folkman, Judah
CC       APPLICANT:  O'Reilly, Michael
CC       TITLE OF INVENTION:  Method of Diagnosing an Angiogenic
CC       TITLE OF INVENTION:  Disease
CC       NUMBER OF SEQUENCES:  6
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Jones & Askew
CC         STREET:  191 Peachtree Street, 37th Floor
CC         CITY:  Atlanta
CC         STATE:  Georgia
CC         COUNTRY:  USA
CC         ZIP:  30303-1769
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette, 3.50
CC         COMPUTER:  MacIntosh
CC         OPERATING SYSTEM:  7.0
CC         SOFTWARE:  Microsoft Word
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/452,260
CC         FILING DATE:  05/26/95
CC         CLASSIFICATION:  514
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  08/248,629
CC         FILING DATE:  04/26/94
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Larry W. Stults, Ph.D.
CC         REGISTRATION NUMBER:  34,025
CC         REFERENCE/DOCKET NUMBER:  05213-0124
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:    404-818-3700
CC         TELEFAX:  404-818-3799
CC      INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  812
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
SQ   SEQUENCE   812 AA; 90846 MW; 3411937 CN;

  Query Match             9.8%;  Score 184;  DB 2;  Length 812;
  Best Local Similarity  27.8%;  Pred. No. 8.64e-07;
  Matches    35;  Conservative   38;  Mismatches  43;  Indels  10;  Gaps   9;

Db     377 CYQSDGQSYRGTSSTTITGKKCQSWAAMFPHRHSKTPENFPDAGLEMNYCRNPDGD-KGP 435
           |:  :|: ||   ::   | :| :|    ::  : ::   : || : :|||||| | :||
Qy      25 CFWDNGHLYREDQTSPAPGLRCLNWLD--AQS-GLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db     436 WCYTT-DPSVR-WEYCNLKRCSETGG-SVVELPTVSQEPS-GPSDSETDCMYGNGKDYRG 491
           ||| : :::|     |:  ||:|| : ::  : |  ||:| ||:  | : ::: ::   :
Qy      81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQ-VFAPANALPA 139

Db     492 KTAVTA 497
           ::  :|
Qy     140 RSEAAA 145


RESULT    5
ID   US-08-429-743-1     STANDARD;     PRT;   812 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08429743
XX
CC    Sequence 1, Application US/08429743
CC    Patent No. 5885795
CC     GENERAL INFORMATION:
CC       APPLICANT:  O'Reilly, Michael
CC       APPLICANT:  Folkman, M. Judah
CC       APPLICANT:  Sim, Kim Lee
CC       APPLICANT:  Cao, Yihai
CC       TITLE OF INVENTION:  Angiostatin and Method of Use
CC       NUMBER OF SEQUENCES:  6
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Jones & Askew
CC         STREET:  191 Peachtree Street, 37th Floor
CC         CITY:  Atlanta
CC         STATE:  Georgia
CC         COUNTRY:  U.S.
CC         ZIP:  30303-1769
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/429,743
CC         FILING DATE:
CC         CLASSIFICATION:  424
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/248,629
CC         FILING DATE:  26-APR-1994
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/326,785
CC         FILING DATE:  20-OCT-1994
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Johnson, James D.
CC         REGISTRATION NUMBER:  31,771
CC         REFERENCE/DOCKET NUMBER:  05213-0122
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  404-818-3700
CC         TELEFAX:  404-818-3799
CC      INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  812 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
CC       ORIGINAL SOURCE:
CC         ORGANISM:  Murine
SQ   SEQUENCE   812 AA; 90846 MW; 3411937 CN;
```

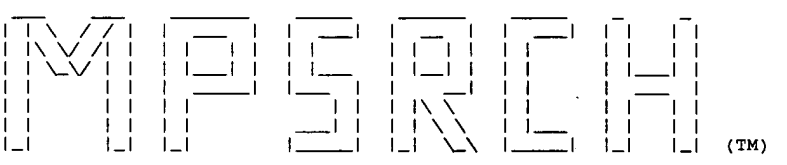
```
  Query Match             9.8%;  Score 184;  DB 2;  Length 812;
  Best Local Similarity  27.8%;  Pred. No. 8.64e-07;
  Matches    35;  Conservative   38;  Mismatches  43;  Indels  10;  Gaps   9;

Db    377 CYQSDGQSYRGTSSTTITGKKCQSWAAMFPHRHSKTPENFPDAGLEMNYCRNPDGD-KGP 435
          |:  :|: ||'  ::   | :| :|    ::  : ::   : || : :|||||| | :||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLD--AQS-GLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db    436 WCYTT-DPSVR-WEYCNLKRCSETGG-SVVELPTVSQEPS-GPSDSETDCMYGNGKDYRG 491
          ||| : :::|     |:  ||:|| : ::  : |  ||:| ||:  | : ::: ::   :
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQ-VFAPANALPA 139

Db    492 KTAVTA 497
          ::  :|
Qy    140 RSEAAA 145


RESULT    6
ID   US-08-451-932-1     STANDARD;     PRT;   812 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08451932
XX
CC    Sequence 1, Application US/08451932
CC    Patent No. 5733876
CC     GENERAL INFORMATION:
CC       APPLICANT:  Folkman, Judah
CC       APPLICANT:  O'Reilly, Michael
CC       TITLE OF INVENTION:  Method of Treating an Angiogenic
CC       TITLE OF INVENTION:  Disease
CC       NUMBER OF SEQUENCES:  6
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Jones & Askew
CC         STREET:  191 Peachtree Street, 37th Floor
CC         CITY:  Atlanta
CC         STATE:  Georgia
CC         COUNTRY:  USA
CC         ZIP:  30303-1769
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette, 3.50
CC         COMPUTER:  MacIntosh
CC         OPERATING SYSTEM:  7.0
CC         SOFTWARE:  Microsoft Word
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/451,932
CC         FILING DATE:  05/26/95
CC         CLASSIFICATION:  514
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  08/248,629
CC         FILING DATE:  04/26/94
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Larry W. Stults, Ph.D.
CC         REGISTRATION NUMBER:  34,025
CC         REFERENCE/DOCKET NUMBER:  05213-0123
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:    404-818-3700
CC         TELEFAX:  404-818-3799
CC     INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  812
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
SQ   SEQUENCE   812 AA; 90846 MW; 3411937 CN;

  Query Match             9.8%;  Score 184;  DB 1;  Length 812;
  Best Local Similarity  27.8%;  Pred. No. 8.64e-07;
  Matches    35;  Conservative   38;  Mismatches  43;  Indels  10;  Gaps   9;

Db    377 CYQSDGQSYRGTSSTTITGKKCQSWAAMFPHRHSKTPENFPDAGLEMNYCRNPDGD-KGP 435
          |:  :|: ||   ::   | :| :|    ::  : ::   : || : :|||||| | :||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLD--AQS-GLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db    436 WCYTT-DPSVR-WEYCNLKRCSETGG-SVVELPTVSQEPS-GPSDSETDCMYGNGKDYRG 491
          ||| : :::|     |:  ||:|| : ::  : |  ||:| ||:  | : ::: ::   :
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQ-VFAPANALPA 139

Db    492 KTAVTA 497
          ::  :|
Qy    140 RSEAAA 145


RESULT    7
ID   US-08-326-785-1     STANDARD;     PRT;   812 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08326785
XX
CC    Sequence 1, Application US/08326785
CC    Patent No. 5792845
CC     GENERAL INFORMATION:
CC       APPLICANT:  Folkman, Judah
CC       APPLICANT:  O'Reilly, Michael
CC       TITLE OF INVENTION:  Angiostatin and Method of Use
CC       NUMBER OF SEQUENCES:  6
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Jones & Askew
CC         STREET:  191 Peachtree Street, 37th Floor
CC         CITY:  Atlanta
CC         STATE:  Georgia
CC         COUNTRY:  USA
CC         ZIP:  30303-1769
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette, 3.50
CC         COMPUTER:  MacIntosh
CC         OPERATING SYSTEM:  7.0
CC         SOFTWARE:  Microsoft Word
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/326,785
CC         FILING DATE:
CC         CLASSIFICATION:  424
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  08/248,629
CC         FILING DATE:  04/26/94
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Larry W. Stults, Ph.D.
CC         REGISTRATION NUMBER:  34,025
CC         REFERENCE/DOCKET NUMBER:  05213-0121
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:    404-818-3700
CC         TELEFAX:  404-818-3799
CC     INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  812
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
SQ   SEQUENCE   812 AA; 90846 MW; 3411937 CN;

  Query Match             9.8%;  Score 184;  DB 2;  Length 812;
  Best Local Similarity  27.8%;  Pred. No. 8.64e-07;
  Matches    35;  Conservative   38;  Mismatches  43;  Indels  10;  Gaps   9;

Db    377 CYQSDGQSYRGTSSTTITGKKCQSWAAMFPHRHSKTPENFPDAGLEMNYCRNPDGD-KGP 435
          |:  :|: ||   ::   | :| :|    ::  : ::   : || : :|||||| | :||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLD--AQS-GLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db    436 WCYTT-DPSVR-WEYCNLKRCSETGG-SVVELPTVSQEPS-GPSDSETDCMYGNGKDYRG 491
          ||| : :::|     |:  ||:|| : ::  : |  ||:| ||:  | : ::: ::   :
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQ-VFAPANALPA 139
```

```
Db    492 KTAVTA 497
          ::  :|
Qy    140 RSEAAA 145


RESULT    8
ID   US-08-248-629A-1     STANDARD;     PRT;   812 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08248629A
XX
CC    Sequence 1, Application US/08248629A
CC    Patent No. 5639725
CC     GENERAL INFORMATION:
CC       APPLICANT:  Folkman, Judah
CC       APPLICANT:  O'Reilly, Michael
CC       TITLE OF INVENTION:  Angiostatin and Method of Use
CC       NUMBER OF SEQUENCES:  6
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Jones & Askew
CC         STREET:  191 Peachtree Street, 37th Floor
CC         CITY:  Atlanta
CC         STATE:  Georgia
CC         COUNTRY:  USA
CC         ZIP:  30303-1769
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette, 3.50
CC         COMPUTER:  MacIntosh
CC         OPERATING SYSTEM:  7.0
CC         SOFTWARE:  Microsoft Word
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/248,629A
CC         FILING DATE:  04/26/94
CC         CLASSIFICATION:  514
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Larry W. Stults, Ph.D.
CC         REGISTRATION NUMBER:  34,025
CC         REFERENCE/DOCKET NUMBER:  05213-0120
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  404-818-3700
CC         TELEFAX:  404-818-3799
CC      INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  812
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
SQ   SEQUENCE   812 AA; 90846 MW; 3411937 CN;

  Query Match            9.8%;  Score 184;  DB 1;  Length 812;
  Best Local Similarity  27.8%;  Pred. No. 8.64e-07;
  Matches    35;  Conservative   38;  Mismatches  43;  Indels  10;  Gaps   9;

Db    377 CYQSDGQSYRGTSSTTITGKKCQSWAAMFPHRHSKTPENFPDAGLEMNYCRNPDGD-KGP 435
          |:  :|: ||   ::   | :| :|    ::  : ::   : || : :|||||| | :||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLD--AQS-GLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db    436 WCYTT-DPSVR-WEYCNLKRCSETGG-SVVELPTVSQEPS-GPSDSETDCMYGNGKDYRG 491
          ||| : :::|     |:  ||:|| : ::  : |  ||:| ||:  | : ::: ::   :
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQ-VFAPANALPA 139

Db    492 KTAVTA 497
          ::  :|
Qy    140 RSEAAA 145


RESULT    9
ID   US-08-148-910-12     STANDARD;     PRT;   655 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 12, Application US/08148910
XX
CC    Sequence 12, Application US/08148910
CC    Patent No. 5466593
CC     GENERAL INFORMATION:
CC       APPLICANT:  Takeshi SHIMOMURA et al.
CC       TITLE OF INVENTION:  No. 5466593el Protein and Gene Encoding Said Protein
CC       NUMBER OF SEQUENCES:  14
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Wenderoth, Lind & Ponack
CC         STREET:  805 Fifteenth Street, N.W., #700
CC         CITY:  Washington
CC         STATE:  D.C.
CC         COUNTRY:  U.S.A.
CC         ZIP:  20005
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette, 5.25 inch,
CC         MEDIUM TYPE:  500 Kb Storage
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  MS-DOS
CC         SOFTWARE:  Wordperfect
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/148,910
CC         FILING DATE:  No. 5466593ember 5, 1993
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Warren M. Cheek, Jr.
CC         REGISTRATION NUMBER:  33,367
CC         REFERENCE/DOCKET NUMBER:
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  202-371-8850
CC         TELEFAX:  202-371-8856
CC         TELEX:
CC      INFORMATION FOR SEQ ID NO:  12:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  655 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       ORIGINAL SOURCE:
CC         ORGANISM:  human
SQ   SEQUENCE   655 AA; 70681 MW; 2184486 CN;

  Query Match            9.6%;  Score 181;  DB 1;  Length 655;
  Best Local Similarity  40.5%;  Pred. No. 1.54e-06;
  Matches    30;  Conservative   19;  Mismatches  17;  Indels   8;  Gaps   4;

Db    286 CFLGNGTGYRGVASTSASGLSCLAWNSDLLYQELHVDSVGAAALLGLGPHAYCRNPDNDE 345
          ||  ||  ||   :::|:|| || |    |  : :  ::::|:: | | |:|||||:|
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW----L--DAQ-SGLASAPVSGAGNHSYCRNPDEDP 77

Db    346 R-PWCYVVKDSALS 358
          | |||||  :::::
Qy     78 RGPWCYVSGEAGVP 91


RESULT   10
ID   US-08-448-937A-12     STANDARD;     PRT;   655 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 12, Application US/08448937A
```

```
XX
CC    Sequence 12, Application US/08448937A
CC    Patent No. 5677164
CC     GENERAL INFORMATION:
CC       APPLICANT:  Takeshi SHIMOMURA et al.
CC       TITLE OF INVENTION:  No. 5677164el Protein and Gene Encoding Said Protein
CC       NUMBER OF SEQUENCES:  14
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Wenderoth, Lind & Ponack
CC         STREET:  805 Fifteenth Street, N.W., #700
CC         CITY:  Washington
CC         STATE:  D.C.
CC         COUNTRY:  U.S.A.
CC         ZIP:  20005
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette, 5.25 inch,
CC         MEDIUM TYPE:  500 Kb Storage
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  MS-DOS
CC         SOFTWARE:  Wordperfect
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/448,937A
CC         FILING DATE:  May 24, 1995
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  08/148,910
CC         FILING DATE:  No. 5677164ember 5, 1993
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Warren M. Cheek, Jr.
CC         REGISTRATION NUMBER:  33,367
CC         REFERENCE/DOCKET NUMBER:
CC       TELECOMMUNICATION INFORMATION:
CC   -     TELEPHONE:  202-371-8850
CC         TELEFAX:  202-371-8856
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  12:
CC       SEQUENCE CHARACTERISTICS:
CC   -     LENGTH:  655 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       ORIGINAL SOURCE:
CC         ORGANISM:  human
SQ   SEQUENCE   655 AA; 70681 MW; 2184486 CN;

  Query Match            9.6%;  Score 181;  DB 1;  Length 655;
  Best Local Similarity  40.5%;  Pred. No. 1.54e-06;
  Matches    30;  Conservative   19;  Mismatches  17;  Indels   8;  Gaps   4;

Db    286 CFLGNGTGYRGVASTSASGLSCLAWNSDLLYQELHVDSVGAAALLGLGPHAYCRNPDNDE 345
          ||  ||  ||   :::|:|| || |    |  : :  ::::|:: | | |:||||||:|
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW----L--DAQ-SGLASAPVSGAGNHSYCRNPDEDP 77

Db    346 R-PWCYVVKDSALS 358
          | |||||  :::::
Qy     78 RGPWCYVSGEAGVP 91


RESULT   11
ID   US-08-622-720A-24     STANDARD;     PRT;   217 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 24, Application US/08622720A
XX
CC    Sequence 24, Application US/08622720A
CC    Patent No. 5814308
CC     GENERAL INFORMATION:
CC       APPLICANT:  Zhang, Ke
CC       TITLE OF INVENTION:  METHODS FOR THE TREATMENT OF
CC       TITLE OF INVENTION:  GASTROINTESTINAL TRACT DISORDERS
CC       NUMBER OF SEQUENCES:  25
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Amgen Inc.
CC         STREET:  1840 Dehavilland Drive
CC         CITY:  Thousand Oaks
CC         STATE:  California
CC         COUNTRY:  U.S.A.
CC         ZIP:  91320
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/622,720A
CC         FILING DATE:
CC         CLASSIFICATION:  424
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Winter, Robert B.
CC         REFERENCE/DOCKET NUMBER:  A-396
CC     INFORMATION FOR SEQ ID NO:  24:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  217 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   217 AA; 24688 MW; 238595 CN;

  Query Match            9.3%;  Score 176;  DB 2;  Length 217;
  Best Local Similarity  32.4%;  Pred. No. 4.04e-06;
  Matches    23;  Conservative   20;  Mismatches  23;  Indels   5;  Gaps   5;

Db     87 CYHGAGELYRGSVSKTRKGIRCQNWSAETPHKPQFKHTSAPHTPLEENFCRNPDGDSHGP 146
          |:   | |||   :    |:|| ||  :: : : :  :: : :  : ::|||||| |::||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW-LDA-Q-SGLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db    147 WCYTT-DPGTP 156
          ||| : ::| |
Qy     81 WCYVSGEAGVP 91


RESULT   12
ID   US-08-290-937B-3     STANDARD;     PRT;   723 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 3, Application US/08290937B
XX
CC    Sequence 3, Application US/08290937B
CC    Patent No. 5648233
CC     GENERAL INFORMATION:
CC       APPLICANT:  YAMAGUCHI, KYOJI
CC       APPLICANT:  SHIMA, NOBUYUKI
CC       APPLICANT:  MURAKAMI, AKIHIKO
CC       APPLICANT:  GOTO, MASAAKI
CC       APPLICANT:  TSUDA, EISUKE
CC       APPLICANT:  MASUNAGA, HIROAKI
CC       APPLICANT:  TAKAHIRA, REIKO
CC       APPLICANT:  OOGAKI, FUMIKO
CC       APPLICANT:  UEDA, MASATSUGU
CC       APPLICANT:  HIGASHIO, KANJI
CC       TITLE OF INVENTION:  MODIFIED TCF
CC       NUMBER OF SEQUENCES:  13
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Testa, Hurwitz & Thibeault
CC         STREET:  125 High St.
CC         CITY:  Boston
CC         STATE:  MA
CC         COUNTRY:  USA
```

```
CC         ZIP:  02110
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/290,937B
CC         FILING DATE:  19-AUG-1994
CC         CLASSIFICATION:  530
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  CAMPBELL, PAULA A.
CC         REGISTRATION NUMBER:  32,503
CC         REFERENCE/DOCKET NUMBER:  FJN-022
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (617) 248-7000
CC         TELEFAX:  (617) 248-7100
CC     INFORMATION FOR SEQ ID NO:  3:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  723 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   723 AA; 82600 MW; 2691692 CN;

  Query Match             8.9%;  Score 167;  DB 1;  Length 723;
  Best Local Similarity  31.5%;  Pred. No. 2.26e-05;
  Matches    28;  Conservative   22;  Mismatches  33;  Indels   6;  Gaps   4;

Db    386 CYRGNGKNYMGNLSQTRSGLTCSMWDKNMEDLHRHIFWEPDASKLNENYCRNPDDDAHGP 445
          |:: ||  |  : :   :|| |  |   ::  :  :   | ::  | :||||||:|::||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW---LDA-QSGLASAPVSGAGNHSYCRNPDEDPRGP 80

Db    446 WCY-TGNPLIP-WDYCPISRCEGDTTPTI 472
          ||| :|:: :|    |   ||   |: ::
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQAL 109


RESULT   13
ID   US-08-290-937B-2     STANDARD;     PRT;   723 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08290937B
XX
CC    Sequence 2, Application US/08290937B
CC    Patent No. 5648233
CC     GENERAL INFORMATION:
CC       APPLICANT:  YAMAGUCHI, KYOJI
CC       APPLICANT:  SHIMA, NOBUYUKI
CC       APPLICANT:  MURAKAMI, AKIHIKO
CC       APPLICANT:  GOTO, MASAAKI
CC       APPLICANT:  TSUDA, EISUKE
CC       APPLICANT:  MASUNAGA, HIROAKI
CC       APPLICANT:  TAKAHIRA, REIKO
CC       APPLICANT:  OOGAKI, FUMIKO
CC       APPLICANT:  UEDA, MASATSUGU
CC       APPLICANT:  HIGASHIO, KANJI
CC       TITLE OF INVENTION:  MODIFIED TCF
CC       NUMBER OF SEQUENCES:  13
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Testa, Hurwitz & Thibeault
CC         STREET:  125 High St.
CC         CITY:  Boston
CC         STATE:  MA
CC         COUNTRY:  USA
CC         ZIP:  02110
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/290,937B
CC         FILING DATE:  19-AUG-1994
CC         CLASSIFICATION:  530
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  CAMPBELL, PAULA A.
CC         REGISTRATION NUMBER:  32,503
CC         REFERENCE/DOCKET NUMBER:  FJN-022
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (617) 248-7000
CC         TELEFAX:  (617) 248-7100
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  723 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   723 AA; 82586 MW; 2690825 CN;

  Query Match             8.9%;  Score 167;  DB 1;  Length 723;
  Best Local Similarity  31.5%;  Pred. No. 2.26e-05;
  Matches    28;  Conservative   22;  Mismatches  33;  Indels   6;  Gaps   4;

Db    386 CYRGNGKNYMGNLSQTRSGLTCSMWDKNMEDLHRHIFWEPDASKLNENYCRNPDDDAHGP 445
          |:: ||  |  : :   :|| |  |   ::  :  :   | ::  | :||||||:|::||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW---LDA-QSGLASAPVSGAGNHSYCRNPDEDPRGP 80

Db    446 WCY-TGNPLIP-WDYCPISRCEGDTTPTI 472
          ||| :|:: :|    |   ||   |: ::
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQAL 109


RESULT   14
ID   US-08-290-937B-1     STANDARD;     PRT;   723 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08290937B
XX
CC    Sequence 1, Application US/08290937B
CC    Patent No. 5648233
CC     GENERAL INFORMATION:
CC       APPLICANT:  YAMAGUCHI, KYOJI
CC       APPLICANT:  SHIMA, NOBUYUKI
CC       APPLICANT:  MURAKAMI, AKIHIKO
CC       APPLICANT:  GOTO, MASAAKI
CC       APPLICANT:  TSUDA, EISUKE
CC       APPLICANT:  MASUNAGA, HIROAKI
CC       APPLICANT:  TAKAHIRA, REIKO
CC       APPLICANT:  OOGAKI, FUMIKO
CC       APPLICANT:  UEDA, MASATSUGU
CC       APPLICANT:  HIGASHIO, KANJI
CC       TITLE OF INVENTION:  MODIFIED TCF
CC       NUMBER OF SEQUENCES:  13
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Testa, Hurwitz & Thibeault
CC         STREET:  125 High St.
CC         CITY:  Boston
CC         STATE:  MA
CC         COUNTRY:  USA
CC         ZIP:  02110
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/290,937B
```

```
CC         FILING DATE:  19-AUG-1994
CC         CLASSIFICATION:  530
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  CAMPBELL, PAULA A.
CC         REGISTRATION NUMBER:  32,503
CC         REFERENCE/DOCKET NUMBER:  FJN-022
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (617) 248-7000
CC         TELEFAX:  (617) 248-7100
CC     INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  723 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   723 AA; 82602 MW; 2699270 CN;

  Query Match             8.9%;  Score 167;  DB 1;  Length 723;
  Best Local Similarity  31.5%;  Pred. No. 2.26e-05;
  Matches    28;  Conservative   22;  Mismatches  33;  Indels   6;  Gaps   4;

Db    386 CYRGNGKNYMGNLSQTRSGLTCSMWDKNMEDLHRHIFWEPDASKLNENYCRNPDDDAHGP 445
          |:: ||  |  : :   :|| |  |   ::  :  :   | ::  | :||||||:|::|| |
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW---LDA-QSGLASAPVSGAGNHSYCRNPDEDPRGP 80

Db    446 WCY-TGNPLIP-WDYCPISRCEGDTTPTI 472
          ||| :|:: :|    |   ||   |: ::
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQAL 109


RESULT   15
ID   US-08-404-643-1     STANDARD;     PRT;   723 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08404643
XX
CC    Sequence 1, Application US/08404643
CC    Patent No. 5658742
CC     GENERAL INFORMATION:
CC       APPLICANT:  HIGASHIO, KANJI
CC       APPLICANT:  SHIMA, NOBUYUKI
CC       APPLICANT:  OOGAKI, FUMIKO
CC       TITLE OF INVENTION:  MONOCLONAL ANTIBODY
CC       NUMBER OF SEQUENCES:  1
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  PATENT ADMINISTRATOR, TESTA, HURWITZ &
CC         ADDRESSEE:  THIBEAULT
CC         STREET:  53 STATE STREET
CC         CITY:  BOSTON
CC         STATE:  MA
CC         COUNTRY:  USA
CC         ZIP:  02109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/404,643
CC         FILING DATE:  15-MAR-1995
CC         CLASSIFICATION:  436
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  CAMPBELL, PAULA A
CC         REGISTRATION NUMBER:  32,503
CC         REFERENCE/DOCKET NUMBER:  FJN-034 (3999/35)
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (617)-248-7000
CC         TELEFAX:  (617)-248-7100
CC     INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  723 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   723 AA; 82602 MW; 2699270 CN;

  Query Match             8.9%;  Score 167;  DB 1;  Length 723;
  Best Local Similarity  31.5%;  Pred. No. 2.26e-05;
  Matches    28;  Conservative   22;  Mismatches  33;  Indels   6;  Gaps   4;

Db    386 CYRGNGKNYMGNLSQTRSGLTCSMWDKNMEDLHRHIFWEPDASKLNENYCRNPDDDAHGP 445
          |:: ||  |  : :   :|| |  |   ::  :  :   | ::  | :||||||:|::|| |
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW---LDA-QSGLASAPVSGAGNHSYCRNPDEDPRGP 80

Db    446 WCY-TGNPLIP-WDYCPISRCEGDTTPTI 472
          ||| :|:: :|    |   ||   |: ::
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQAL 109


Search completed: Fri Sep 17 18:29:44 1999
Job time : 35 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 17 18:25:49 1999; MasPar time 12.81 Seconds
822.688 Million cell updates/sec
Tabular output not generated.

Title: >US-09-084-491A-2
Description: (1-263) from US09084491A.pep
Perfect Score: 1883
Sequence: 1 MLLAWVQAFLVSNMLLAEAY.....PVDPQEGSTPLMGQAGTPGA 263

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 45.508; Variance 88.911; scale 0.512

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	204	10.8	169	2	A40522	plasmin (EC 3.4.21.7)	2.52e-17
2	184	9.8	812	1	PLMS	plasmin (EC 3.4.21.7)	7.26e-14
3	183	9.7	4548	1	S00657	apoprotein(a) (EC 3.4	1.07e-13
4	181	9.6	655	1	A46688	hepatocyte growth fac	2.35e-13
5	178	9.5	716	1	JC5061	macrophage-stimulatin	7.55e-13
6	177	9.4	685	2	A48289	neurotrophic receptor	1.11e-12
7	177	9.4	716	1	A40332	macrophage-stimulatin	1.11e-12
8	176	9.3	810	2	I46260	plasmin (EC 3.4.21.7)	1.64e-12
9	173	9.2	810	2	B30848	plasmin (EC 3.4.21.7)	5.22e-12
10	171	9.1	790	1	PLPG	plasmin (EC 3.4.21.7)	1.13e-11
11	172	9.1	812	1	PLBO	plasmin (EC 3.4.21.7)	7.67e-12
12	168	8.9	728	1	A60185	hepatocyte growth fac	3.54e-11
13	167	8.9	728	1	JH0579	hepatocyte growth fac	5.19e-11
14	165	8.8	622	1	TBHU	thrombin (EC 3.4.21.5	1.11e-10
15	166	8.8	711	1	A47136	macrophage-stimulatin	7.58e-11
16	165	8.8	728	1	A35644	hepatocyte growth fac	1.11e-10
17	164	8.7	625	1	TBBO	thrombin (EC 3.4.21.5	1.62e-10
18	163	8.7	1420	2	A32869	apolipoprotein(a) (EC	2.36e-10
19	161	8.6	810	1	PLHU	plasmin (EC 3.4.21.7)	5.01e-10
20	158	8.4	603	2	S28941	coagulation factor XI	1.54e-09
21	156	8.3	89	2	A60140	plasmin (EC 3.4.21.7)	3.24e-09
22	156	8.3	431	2	JS0599	t-plasminogen activat	3.24e-09
23	156	8.3	477	2	JS0598	t-plasminogen activat	3.24e-09
24	156	8.3	477	2	A34369	t-plasminogen activat	3.24e-09
25	152	8.1	120	2	E61545	plasmin (EC 3.4.21.7)	1.42e-08
26	152	8.1	433	1	UKBAY	u-plasminogen activat	1.42e-08
27	152	8.1	434	1	A35005	u-plasminogen activat	1.42e-08
28	151	8.0	432	1	S18932	u-plasminogen activat	2.05e-08
29	150	8.0	442	1	UKPG	u-plasminogen activat	2.96e-08
30	150	8.0	562	1	UKHUT	t-plasminogen activat	2.96e-08
31	151	8.0	943	2	B45082	neurotrophic receptor	2.05e-08
32	149	7.9	593	2	S45281	coagulation factor XI	4.26e-08
33	147	7.8	123	2	C61545	plasmin (EC 3.4.21.7)	8.83e-08
34	147	7.8	394	2	JS0600	t-plasminogen activat	8.83e-08
35	146	7.8	411	2	S65783	plasminogen activator	1.27e-07
36	146	7.8	431	1	UKHU	u-plasminogen activat.	1.27e-07
37	147	7.8	433	1	UKMS	u-plasminogen activat	8.83e-08
38	147	7.8	455	2	A61545	plasmin (EC 3.4.21.7)	8.83e-08
39	147	7.8	460	2	B61545	plasmin (EC 3.4.21.7)	8.83e-08
40	146	7.8	617	2	S10511	thrombin (EC 3.4.21.5	1.27e-07
41	144	7.6	618	2	A35827	thrombin (EC 3.4.21.5	2.61e-07
42	141	7.5	433	1	JN0560	u-plasminogen activat	7.65e-07
43	141	7.5	559	2	A29941	t-plasminogen activat	7.65e-07
44	142	7.5	710	1	I51283	hepatocyte growth fac	5.35e-07
45	138	7.3	291	2	I38098	t-plasminogen activat	2.22e-06

ALIGNMENTS

```
RESULT    1
ENTRY             A40522     #type fragment
TITLE             plasmin (EC 3.4.21.7) precursor - rat (fragment)
ORGANISM          #formal_name Rattus norvegicus #common_name Norway rat
DATE              28-Feb-1992 #sequence_revision 17-Apr-1993 #text_change
                    08-Sep-1997
ACCESSIONS        A40522
REFERENCE         A40522
   #authors       Kanalas, J.J.; Makker, S.P.
   #journal       J. Biol. Chem. (1991) 266:10825-10829
   #title         Identification of the rat Heymann nephritis autoantigen
                    (GP330) as a receptor site for plasminogen.
   #cross-references MUID:91250378
   #accession     A40522
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-169 ##label KAN
      ##cross-references GB:M62832; NID:g206215; PID:g554488
      ##note           the authors translated the codon TCT for residue 76 as
                        Ala
CLASSIFICATION    #superfamily plasmin; kringle homology; plasminogen-related
                    protein precursor homology; trypsin homology
KEYWORDS          fibrinolysis; glycoprotein; hydrolase; kringle; serine
                    proteinase
FEATURE
   34-112              #domain kringle homology #label KR1\
   34-112,55-95,
   83-107              #disulfide_bonds #status predicted
SUMMARY           #length 169  #checksum 4101

  Query Match           10.8%;  Score 204;  DB 2;  Length 169;
  Best Local Similarity  26.2%;  Pred. No. 2.52e-17;
  Matches    33;  Conservative   44;  Mismatches  40;  Indels   9;  Gaps   8;

Db      34 CYQGNGKSYRGTSSTTNTGKKCQSWVSMTPHSHSKTPANFPDSGLEMNYCRNPDNDQRGP 93
           |:  ||  ||   ::   | :| :|  : ::| : ::|  : :| : :||||||:| |||
Qy      25 CFWDNGHLYREDQTSPAPGLRCLNW--LDAQS-GLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db      94 WCFTT-DPSVR-WEYCNLKRCSETGG-GVAE-SAIVPQVPSAPGTSETDCMYGNGKEYRG 149
           ||: : :::|     |:  ||:|| : :::  :: : : : :||: | : ::: ::   :
Qy      81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQ-VFAPANALPA 139

Db     150 KTAVTA 155
           ::  :|
Qy     140 RSEAAA 145
```

```
RESULT    2
ENTRY             PLMS       #type complete
TITLE             plasmin (EC 3.4.21.7) precursor - mouse
CONTAINS          angiostatin; plasminogen
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              20-Sep-1991 #sequence_revision 01-Nov-1996 #text_change
                    08-May-1998
ACCESSIONS        A38514; S48202; S48203
REFERENCE         A38514
   #authors       Degen, S.J.F.; Bell, S.M.; Schaefer, L.A.; Elliott, R.W.
   #journal       Genomics (1990) 8:49-61
   #title         Characterization of the cDNA coding for mouse plasminogen and
                    localization of the gene to mouse chromosome 17.
   #cross-references MUID:91184812
   #accession     A38514
      ##molecule_type mRNA
      ##residues       1-812 ##label DEG
      ##cross-references GB:J04766; NID:g200402; PID:g200403
REFERENCE         S48202
   #authors       Lijnen, H.R.; van Hoef, B.; Beelen, V.; Collen, D.
   #journal       Eur. J. Biochem. (1994) 224:863-871
   #title         Characterization of the murine plasma fibrinolytic system.
   #cross-references MUID:95010076
   #accession     S48202
      ##molecule_type protein
      ##residues       20-25 ##label LIJ
   #accession     S48203
      ##molecule_type protein
      ##residues       22-27 ##label LI2
COMMENT     Plasminogen is synthesized by the kidney and is present in plasma
              and many other extracellular fluids.
COMMENT     Plasminogen is converted into plasmin by plasminogen activators,
      -       both plasminogen and its activator being bound to fibrin. Plasmin
              is inactivated by alpha-2-antiplasmin (see PIR:S47217)
              immediately after dissociation from the clot. In the presence of
              the inhibitor, the activation involves only cleavage after
              Arg-581, resulting in two chains connected by two disulfide
      -       bonds. Without the inhibitor, the activation involves also
              removal of the activation peptide.
COMMENT     Stromelysin 1 (see PIR:KCMSS1) acts on plasminogen to produce
              angiostatin. Together with endostatin (see PIR:A56101,
              PIR:B56101), angiostatin acts to inhibit angiogenesis, and so may
              be useful in treating solid tumors.
FUNCTION
   #description  dissolves the fibrin of blood clots; acts as a proteolytic
                   factor in a variety of processes including embryonic
                   development, tissue remodeling and tumor invasion; in
                   ovulation it weakens the walls of the graafian follicle;
                   also activates the urokinase-type plasminogen activator
   #pathway      fibrinolysis
CLASSIFICATION   #superfamily plasmin; kringle homology; plasminogen-related
                   protein precursor homology; trypsin homology
KEYWORDS         angiogenesis inhibitor; blood; duplication; fibrinolysis;
                   glycoprotein; hydrolase; kidney; kringle; plasma; serine
                   proteinase; zymogen

FEATURE
   1-96               #domain plasminogen-related protein precursor homology
                        #label PLPH\
   1-19               #domain signal sequence #status predicted #label SIG\
   20-812             #product plasminogen #status predicted #label PRO\
   20-96              #domain activation peptide #status predicted #label APT\
   79-466             #product angiostatin #status predicted #label AST\
   97-581,582-812     #product plasmin #status predicted #label MAT\
   97-581             #domain chain A #status predicted #label ACH\
   103-181            #domain kringle homology #label KR1\
   185-262            #domain kringle homology #label KR2\
   275-352            #domain kringle homology #label KR3\
   377-454            #domain kringle homology #label KR4\
   481-560            #domain kringle homology #label KR5\
   582-812            #domain chain B #status predicted #label BCH\
   582-805            #domain trypsin homology #label TRY\
   49-73,53-61,
   103-181,124-164,
   152-176,185-262,
   188-316,206-245,
   234-257,275-352,
   296-335,324-347,
   377-454,398-437,
   426-449,481-560,
   502-543,531-555,
   568-687,578-586,
   609-625,701-768,
   731-747,758-786    #disulfide_bonds #status predicted\
   78-79              #cleavage_site Glu-Asn (stromelysin 1) #status
                        predicted\
   136,308            #binding_site carbohydrate (Asn) (covalent) #status
                        predicted\
   466-467            #cleavage_site Thr-Val (stromelysin 1) #status
                        predicted\
   581-582            #cleavage_site Arg-Val (plasminogen activator) #status
                        experimental\
   624,667,762        #active_site His, Asp, Ser #status predicted
SUMMARY          #length 812  #molecular-weight 90846  #checksum 237

  Query Match    .        9.8%;  Score 184;  DB 1;  Length 812;
  Best Local Similarity  27.8%;  Pred. No. 7.26e-14;
  Matches    35;  Conservative   38;  Mismatches  43;  Indels  10;  Gaps   9;

Db    377 CYQSDGQSYRGTSSTTITGKKCQSWAAMFPHRHSKTPENFPDAGLEMNYCRNPDGD-KGP 435
          |:  :|: ||   ::   | :| :|    ::  : ::   : || : :|||||| | :||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLD--AQS-GLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db    436 WCYTT-DPSVR-WEYCNLKRCSETGG-SVVELPTVSQEPS-GPSDSETDCMYGNGKDYRG 491
          ||| : :::|     |:  ||:|| : ::  : |  ||:| ||:  | : ::: ::   :
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQ-VFAPANALPA 139

Db    492 KTAVTA 497
          ::  :|
Qy    140 RSEAAA 145


RESULT    3
ENTRY             S00657     #type complete
TITLE             apoprotein(a) (EC 3.4.21.-) precursor - human
ALTERNATE_NAMES   apolipoprotein(a); lipoprotein(a) chain apo(a)
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change
                    24-Oct-1997
ACCESSIONS        S00657; A28017; A47277; I60906; A47233; I52415; I65286
REFERENCE         S00657
   #authors       McLean, J.W.; Tomlinson, J.E.; Kuang, W.J.; Eaton, D.L.;
                    Chen, E.Y.; Fless, G.M.; Scanu, A.M.; Lawn, R.M.
   #journal       Nature (1987) 330:132-137
   #title         cDNA sequence of human apolipoprotein(a) is homologous to
                    plasminogen.
   #cross-references MUID:88039109
   #accession     S00657
      ##molecule_type mRNA
      ##residues       1-4548 ##label MCL
      ##cross-references GB:X06290; EMBL:X06696; NID:g28619; PID:g28620
REFERENCE         A28017
   #authors       Eaton, D.L.; Fless, G.M.; Kohr, W.J.; McLean, J.W.; Xu, Q.T.;
                    Miller, C.G.; Lawn, R.M.; Scanu, A.M.
   #journal       Proc. Natl. Acad. Sci. U.S.A. (1987) 84:3224-3228
   #title         Partial amino acid sequence of apolipoprotein(a) shows that
                    it is homologous to plasminogen.
   #cross-references MUID:87204109
   #accession     A28017
      ##molecule_type protein
      ##residues       20-21,'P',23-34;177-179,'N',181-186,'T',188-196,'DKG',
                        200;292-314,'W',316-318;4201-4202,'X',4204-4207,'LL',
                        4210-4212,'P',4214-4218,'X',4220-4232;4323-4334,'LTP';
                        4382-4386,'X',4388-4393,'PX',4396-4401 ##label EAT
REFERENCE         A47277
   #authors       Wade, D.P.; Clarke, J.G.; Lindahl, G.E.; Liu, A.C.; Zysow,
                    B.R.; Meer, K.; Schwartz, K.; Lawn, R.M.
```

```
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1993) 90:1369-1373
   #title        5' control regions of the apolipoprotein(a) gene and members
                   of the related plasminogen gene family.
   #cross-references MUID:93165698
   #accession    A47277
      ##status        preliminary; translation not shown; translated from
                        GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues      1-16 ##label RES
      ##cross-references GB:L07899; NID:g967973; PID:g967974
REFERENCE        A47233
   #authors      Malgaretti, N.; Acquati, F.; Magnaghi, P.; Bruno, L.;
                   Pontoglio, M.; Rocchi, M.; Saccone, S.; Della Valle, G.;
                   D'Urso, M.; LePaslier, D.; Ottolenghi, S.; Taramelli, R.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1992) 89:11584-11588
   #title        Characterization by yeast artificial chromosome cloning of
                   the linked apolipoprotein(a) and plasminogen genes and
                   identification of the apolipoprotein(a) 5' flanking region.
   #cross-references MUID:93087573
   #accession    I60906
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues      1-16 ##label RE2
      ##cross-references GB:M90078; NID:g178786; PID:g553188
      ##note          apo(a) gene 1 (nomenclature of reference I52415)
   #accession    A47233
      ##status        preliminary; translation not shown; translated from
                        GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues      1-16 ##label RE5
      ##cross-references GB:M90079; NID:g178784; PID:g553187
REFERENCE        I52415
   #authors      Ichinose, A.
   #journal      Biochemistry (1992) 31:3113-3118
   #title        Multiple members of the plasminogen-apolipoprotein(a) gene
                   family associated with thrombosis.
   #cross-references MUID:92207924
   #accession    I52415
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues      1-16 ##label RE3
      ##cross-references GB:M86877; NID:g178780; PID:g553185
      ##note          apo(a) gene 1 (nomenclature of reference I52415)
   #accession    I65286
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues      1-16 ##label RE4
      ##cross-references GB:M86878; NID:g178782; PID:g553186
GENETICS
   #gene         GDB:LPA
      ##cross-references GDB:120699; OMIM:152200
   #map_position 6q26-6q27
   #note         several genes closely linked on chromosome 6 are identical in
                   the first coding exon; the products of any one expressed
                   locus are highly variable in human populations because of
                   variable numbers of kringle repeats
CLASSIFICATION   #superfamily apolipoprotein(a); kringle homology; trypsin
                   homology
KEYWORDS         hydrolase; kringle; lipid binding; lipoprotein; serine
                   proteinase

FEATURE
   1-19              #domain signal sequence #status predicted #label SIG\
   20-4548           #product apolipoprotein(a) #status experimental #label
                       MAT\
   28-105            #domain kringle homology #label KR1\
   142-219           #domain kringle homology #label KR2\
   256-333           #domain kringle homology #label KR3\
   370-447           #domain kringle homology #label KR4\
   484-561           #domain kringle homology #label KR5\
   598-675           #domain kringle homology #label KR6\
   712-789           #domain kringle homology #label KR7\
   826-903           #domain kringle homology #label KR8\
   940-1017          #domain kringle homology #label KR9\
   1054-1131         #domain kringle homology #label KR10\
   1168-1245         #domain kringle homology #label KR11\
   1282-1359         #domain kringle homology #label KR12\
   1396-1473         #domain kringle homology #label KR13\
   1510-1587         #domain kringle homology #label KR14\
   1624-1701         #domain kringle homology #label KR15\
   1738-1815         #domain kringle homology #label KR16\
   1852-1929         #domain kringle homology #label KR17\
   1966-2043         #domain kringle homology #label KR18\
   2080-2157         #domain kringle homology #label KR19\
   2194-2271         #domain kringle homology #label KR20\
   2308-2385         #domain kringle homology #label KR21\
   2422-2499         #domain kringle homology #label KR22\
   2536-2613         #domain kringle homology #label KR23\
   2650-2727         #domain kringle homology #label KR24\
   2764-2841         #domain kringle homology #label KR25\
   2878-2955         #domain kringle homology #label KR26\
   2992-3069         #domain kringle homology #label KR27\
   3106-3183         #domain kringle homology #label KR28\
   3220-3297         #domain kringle homology #label KR29\
   3334-3411         #domain kringle homology #label KR30\
   3448-3525         #domain kringle homology #label KR31\
   3562-3639         #domain kringle homology #label KR32\
   3676-3753         #domain kringle homology #label KR33\
   3782-3859         #domain kringle homology #label KR34\
   3896-3973         #domain kringle homology #label KR35\
   4010-4087         #domain kringle homology #label KR36\
   4124-4201         #domain kringle homology #label KR37\
   4228-4307         #domain kringle homology #label KR38\
   4328-4541         #domain trypsin homology #label TRY
SUMMARY          #length 4548  #molecular-weight 501316  #checksum 1998

  Query Match             9.7%;  Score 183;  DB 1;  Length 4548;
  Best Local Similarity  27.5%;  Pred. No. 1.07e-13;
  Matches    36;  Conservative   34;  Mismatches  52;  Indels   9;  Gaps   8;

Db   3539 ILAPSLEAFF-EQALTEETPGVQDCYYHYGQSYRGTYSTTVTGRTCQAWSSMTPHQHSRT 3597
            :| : ::||:    |  |: |   |:   |: ||   ::   |  |  |  : : | : :
Qy      1 MLLAWVQAFLVSNMLLAEAYGSGGCFWDNGHLYREDQTSPAPGLRCLNW--LDA-QSGLA 57

Db   3598 PENYPNAGLTRNYCRNPDAEIR-PWCYTM-DPSVR-WEYCNLTQCLVTESSVLATLTV-V 3653
           :   : ||  ::|||||| : | ||||   :::|     |:  :|  | |  |:::|  :
Qy     58 SAPVSGAG-NHSYCRNPDEDPRGPWCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEI 116

Db   3654 PDPSTEASSEE 3664
             ::|  ::::|
Qy    117 QEASEGPGADE 127


RESULT    4
ENTRY             A46688     #type complete
TITLE             hepatocyte growth factor activator (EC 3.4.21.-) precursor -
                    human
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              21-Sep-1993 #sequence_revision 25-Aug-1995 #text_change
                    07-Aug-1998
ACCESSIONS        A46688
REFERENCE         A46688
   #authors       Miyazawa, K.; Shimomura, T.; Kitamura, A.; Kondo, J.;
                    Morimoto, Y.; Kitamura, N.
   #journal       J. Biol. Chem. (1993) 268:10024-10028
   #title         Molecular cloning and sequence analysis of the cDNA for a
                    human serine protease reponsible for activation of
                    hepatocyte growth factor. Structural similarity of the
                    protease precursor to blood coagulation factor XII.
   #cross-references MUID:93252878
   #accession    A46688
      ##molecule_type mRNA
      ##residues      1-655 ##label MIY
      ##cross-references DDBJ:D14012; NID:g219680; PID:g219681
      ##experimental_source liver (mRNA); serum (protein)
      ##note          sequence extracted from NCBI backbone (NCBIN:131227,
                       NCBIP:131228)
```

```
      ##note          parts of the sequence, including the amino ends of the
                        heavy and light chains, confirmed by protein
                        sequencing
FUNCTION
   #description  activates hepatocyte growth factor by specific proteolytic
                   cleavage
   #pathway      tissue repair and regeneration
CLASSIFICATION   #superfamily coagulation factor XII; EGF homology;
                   fibronectin type I repeat homology; fibronectin type II
                   repeat homology; kringle homology; trypsin homology
KEYWORDS         glycoprotein; hydrolase; kringle; liver; plasma; serine
                   proteinase
FEATURE
   1-34              #domain signal sequence #status predicted #label SIG\
   108-148           #domain fibronectin type II repeat homology #label FN2\
   164-197           #domain EGF homology #label EGF1\
   202-237           #domain fibronectin type I repeat homology #label FN1\
   245-278           #domain EGF homology #label EGF2\
   286-367           #domain kringle homology #label KGL\
   373-407           #product hepatocyte growth factor activator light chain
                       #status experimental #label LCH\
   408-655           #product hepatocyte growth factor activator heavy chain
                       #status experimental #label HCH\
   408-641           #domain trypsin homology #label TRY\
   40,48,290,468,492,
   546               #binding_site carbohydrate (Asn) (covalent) #status
                       predicted\
   164-175,169-186,
   188-197,202-230,
   228-237,245-256,
   250-267,269-278,
   286-367,307-349,
   338-362,394-521,
   432-448,440-510,
   535-604,567-583,
   594-622           #disulfide_bonds #status predicted\
   447,497,598       #active_site His, Asp, Ser #status predicted
SUMMARY          #length 655  #molecular-weight 70681  #checksum 645

  Query Match             9.6%;  Score 181;  DB 1;  Length 655;
  Best Local Similarity  40.5%;  Pred. No. 2.35e-13;
  Matches    30;  Conservative   19;  Mismatches  17;  Indels   8;  Gaps   4;

Db    286 CFLGNGTGYRGVASTSASGLSCLAWNSDLLYQELHVDSVGAAALLGLGPHAYCRNPDNDE 345
          ||  ||  ||   :::|:|| || |    |  : :  ::::|:: | | |:||||||:|
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW----L--DAQ-SGLASAPVSGAGNHSYCRNPDEDP 77

Db    346 R-PWCYVVKDSALS 358
          | |||||  :::::
Qy     78 RGPWCYVSGEAGVP 91


RESULT    5
ENTRY            JC5061     #type complete
TITLE            macrophage-stimulating protein 1 precursor - rat
ORGANISM         #formal_name Rattus norvegicus #common_name Norway rat
DATE             31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change
                   13-Sep-1997
ACCESSIONS       JC5061
REFERENCE        JC5061
   #authors      Ohshiro, K.; Iwama, A.; Matsuno, K.; Ezaki, T.; Sakamoto, O.;
                   Hamaguchi, I.; Takasu, N.; Suda, T.
   #journal      Biochem. Biophys. Res. Commun. (1996) 227:273-280
   #title        Molecular cloning of rat macrophage-stimulating protein and
                   its involvement in the male reproductive system.
   #cross-references MUID:97011126
   #accession    JC5061
      ##molecule_type mRNA
      ##residues      1-716 ##label OHS
      ##cross-references EMBL:X95096; NID:g1669718; PID:e223604; PID:g1669719
COMPLEX          disulfide-bonded heterodimer of chains derived from the same
                   precursor
CLASSIFICATION   #superfamily hepatocyte growth factor; kringle homology;
                   trypsin homology
KEYWORDS         duplication; glycoprotein; growth factor; kringle
FEATURE
   1-31              #domain signal sequence #status predicted #label SIG\
   32-488,489-716    #product macrophage-stimulating protein 1 #status
                       predicted #label MAT\
   32-488            #domain macrophage-stimulating protein 1 alpha chain
                       #status predicted #label ACH\
   110-186           #domain kringle homology #label KRI1\
   191-268           #domain kringle homology #label KRI2\
   292-370           #domain kringle homology #label KRI3\
   379-457           #domain kringle homology #label KRI4\
   489-716           #domain macrophage-stimulating protein 1 beta chain
                       #status predicted #label BCH\
   489-709           #domain trypsin homology #label TRY\
   72,305,620        #binding_site carbohydrate (Asn) (covalent) #status
                       predicted
SUMMARY          #length 716  #molecular-weight 80733  #checksum 9217

  Query Match             9.5%;  Score 178;  DB 1;  Length 716;
  Best Local Similarity  41.3%;  Pred. No. 7.55e-13;
  Matches    26;  Conservative    9;  Mismatches  26;  Indels   2;  Gaps   2;

Db    110 CIMDNGASYRGTVARTADGLPCQAWSRRFPNDHKYTPTPKNGLEENFCRNPDGDPRGPWC 169
          |: |||  ||   :  | || |  |     :    :| :  | : ::||||| |||||||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLDA-QSGLASAPVSGAG-NHSYCRNPDEDPRGPWC 82

Db    170 YTT 172
          | :
Qy     83 YVS 85


RESULT    6
ENTRY            A48289     #type complete
TITLE            neurotrophic receptor ror precursor - fruit fly (Drosophila
                   melanogaster)
ALTERNATE_NAMES  trk-related receptor
CONTAINS         protein-tyrosine kinase (EC 2.7.1.112)
ORGANISM         #formal_name Drosophila melanogaster
DATE             27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change
                   17-Mar-1999
ACCESSIONS       A48289
REFERENCE        A48289
   #authors      Wilson, C.; Goberdhan, D.C.I.; Steller, H.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1993) 90:7109-7113
   #title        Dror, a potential neurotrophic receptor gene, encodes a
                   Drosophila homolog of the vertebrate Ror family of
                   Trk-related receptor tyrosine kinases.
   #cross-references MUID:93348222
   #accession    A48289
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-685 ##label WIL
      ##cross-references GB:L20297; NID:g348103; PID:g348104
GENETICS
   #gene         FlyBase:bsk
      ##cross-references FlyBase:FBgn0000229
CLASSIFICATION   #superfamily Drosophila neurotrophic receptor ror; kringle
                   homology; protein kinase homology
KEYWORDS         ATP; glycoprotein; kringle; phosphotransferase; transmembrane
                   protein; tyrosine-specific protein kinase
FEATURE
   237-310           #domain kringle homology #label KR1\
   314-338           #domain transmembrane #status predicted #label TM1\
   408-677           #domain protein kinase homology #label KIN\
   416-424           #region protein kinase ATP-binding motif\
   45,63,129,144,250 #binding_site carbohydrate (Asn) (covalent) #status
                       predicted
SUMMARY          #length 685  #molecular-weight 78142  #checksum 8512

  Query Match             9.4%;  Score 177;  DB 2;  Length 685;
  Best Local Similarity  27.4%;  Pred. No. 1.11e-12;
  Matches    26;  Conservative   20;  Mismatches  46;  Indels   3;  Gaps   3;
```

```
Db    234 TENCYWEDGSTYRGVANVSASGKPCLRWSWLMKEISDFP-ELIG-QNYCRNPGSVENSPW 291
          :  |:|::|  ||     :|:|  || |      ::  |    | ::|||||     :||
Qy     22 SGGCFWDNGHLYREDQTSPAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRGPW 81

Db    292 CFVDSSRE-RIIELCDIPKCADKIWIAIVGTTAAI 325
          |:| :         |:  :|::    |: : |: |
Qy     82 CYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEI 116


RESULT    7
ENTRY             A40332     #type complete
TITLE             macrophage-stimulating protein 1 precursor - mouse
ALTERNATE_NAMES   hepatocyte growth factor-like protein
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change
                    11-Sep-1998
ACCESSIONS        A40332; B40332
REFERENCE         A40332
   #authors       Degen, S.J.F.; Stuart, L.A.; Han, S.; Jamison, C.S.
   #journal       Biochemistry (1991) 30:9781-9791
   #title         Characterization of the mouse cDNA and gene coding for a
                    hepatocyte growth factor-like protein: expression during
                    development.
   #cross-references MUID:92002017
   #accession     A40332
      ##molecule_type DNA
      ##residues       1-716 ##label DEG
      ##cross-references GB:M74180; NID:g193831; PID:g193832
   #accession     B40332
      ##molecule_type mRNA
      ##residues       1-18,'P',20-716 ##label DEG2
      ##cross-references GB:M74181; NID:g193833; PID:g193834
GENETICS
   #introns       18/1; 67/2; 105/1; 143/2; 189/1; 229/2; 269/1; 334/2; 378/1;
                    412/2; 458/1; 470/1; 506/2; 532/2; 581/2; 617/1; 663/3
COMPLEX           disulfide-bonded heterodimer of chains derived from the same
                    precursor
CLASSIFICATION    #superfamily hepatocyte growth factor; kringle homology;
                    trypsin homology
KEYWORDS          duplication; glycoprotein; growth factor; kringle
FEATURE
   1-31               #domain signal sequence #status predicted #label SIG\
   19-488,489-716     #product macrophage-stimulating protein 1 #status
                        experimental #label MAT\
   19-483             #domain alpha chain #status experimental #label ACH\
   110-186            #domain kringle homology #label KR1\
   191-268            #domain kringle homology #label KR2\
   292-370            #domain kringle homology #label KR3\
   379-457            #domain kringle homology #label KR4\
   484-711            #domain beta chain #status experimental #label BCH\
   489-709            #domain trypsin homology #label TRY\
   72,173,305,620     #binding_site carbohydrate (Asn) (covalent) #status
                        predicted
SUMMARY           #length 716  #molecular-weight 80619  #checksum 7627

  Query Match             9.4%;  Score 177;  DB 1;  Length 716;
  Best Local Similarity  41.3%;  Pred. No. 1.11e-12;
  Matches    26;  Conservative    9;  Mismatches  26;  Indels   2;  Gaps   2;

Db    110 CIMDNGVSYRGTVARTAGGLPCQAWSRRFPNDHKYTPTPKNGLEENFCRNPDGDPRGPWC 169
          |: |||  ||   :  | || |  |     :    :| :  | : ::|||||  |||||||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLDA-QSGLASAPVSGAG-NHSYCRNPDEDPRGPWC 82

Db    170 YTT 172
          | :
Qy     83 YVS 85


RESULT    8
ENTRY             I46260     #type complete
TITLE             plasmin (EC 3.4.21.7) precursor - western European hedgehog
ORGANISM          #formal_name Erinaceus europaeus #common_name western
                    European hedgehog
DATE              21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change
                    17-Mar-1999
ACCESSIONS        I46260
REFERENCE         I46259
   #authors       Lawn, R.M.; Boonmark, N.W.; Schwartz, K.; Lindahl, G.E.;
                    Wade, D.P.; Byrne, C.D.; Fong, K.J.; Meer, K.; Patthy, L.
   #journal       J. Biol. Chem. (1995) 270:24004-24009
   #title         The recurring evolution of Lp(a): Insights from cloning of
                    hedgehog apolipoprotein(a).
   #cross-references MUID:96025778
   #accession     I46260
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      1-810 ##label LAW
      ##cross-references EMBL:U33171; NID:g1046360; PID:g1046361
CLASSIFICATION    #superfamily plasmin; kringle homology; plasminogen-related
                    protein precursor homology; trypsin homology
KEYWORDS          hydrolase; serine proteinase
FEATURE
   1-96               #domain plasminogen-related protein precursor homology
                        #label PLPH\
   103-181            #domain kringle homology #label KR1\
   185-262            #domain kringle homology #label KR2\
   379-456            #domain kringle homology #label KR4\
   582-803            #domain trypsin homology #label TRY
SUMMARY           #length 810  #molecular-weight 90902  #checksum 2484

  Query Match             9.3%;  Score 176;  DB 2;  Length 810;
  Best Local Similarity  32.9%;  Pred. No. 1.64e-12;
  Matches    26;  Conservative   20;  Mismatches  27;  Indels   6;  Gaps   6;

Db    107 NGKYYRGTVSKTKTGLTCQKWSAETPHKPRFSPDENPSEGLDQNYCRNPDNDPKGPWCYT 166
          ||  ||   :    || | :|  :: :   :::   :: | :::||||||:||:|||||
Qy     29 NGHLYREDQTSPAPGLRCLNW-LDA-QSG-LASAPVSGAG-NHSYCRNPDEDPRGPWCYV 84

Db    167 M-DPEVRYEY-CEIIQCED 183
          :: |     || ::| :
Qy     85 SGEAGVPEKRPCEDLRCPE 103


RESULT    9
ENTRY             B30848     #type complete
TITLE             plasmin (EC 3.4.21.7) precursor - rhesus macaque
ORGANISM          #formal_name Macaca mulatta #common_name rhesus macaque
DATE              31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change
                    08-Sep-1997
ACCESSIONS        B32869; B30848
REFERENCE         A32869
   #authors       Tomlinson, J.E.; McLean, J.W.; Lawn, R.M.
   #journal       J. Biol. Chem. (1989) 264:5957-5965
   #title         Rhesus monkey apolipoprotein(a). Sequence, evolution, and
                    sites of synthesis.
   #cross-references MUID:89174660
   #accession     B32869
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-810 ##label TOM
      ##cross-references GB:J04697; NID:g342272; PID:g342273
CLASSIFICATION    #superfamily plasmin; kringle homology; plasminogen-related
                    protein precursor homology; trypsin homology
KEYWORDS          fibrinolysis; glycoprotein; hydrolase; kringle; serine
                    proteinase
FEATURE
   1-96               #domain plasminogen-related protein precursor homology
                        #label PLPH\
   1-9                #domain signal sequence #status predicted #label SIG\
   103-181            #domain kringle homology #label KR1\
   185-262            #domain kringle homology #label KR2\
   275-352            #domain kringle homology #label KR3\
   377-454            #domain kringle homology #label KR4\
   481-560            #domain kringle homology #label KR5\
   581-803            #domain trypsin homology #label TRY\
```

```
   49-73,53-61,
   103-181,124-164,
   152-176,185-262,
   188-316,206-245,
   234-257,275-352,
   296-335,324-347,
   377-454,398-437,
   426-449,481-560,
   502-543,531-555,
   567-685,577-585,
   607-623,699-766,
   729-745,756-784    #disulfide_bonds #status predicted\
   622,665,760        #active_site His, Asp, Ser #status predicted
SUMMARY          #length 810  #molecular-weight 90255  #checksum 5081

  Query Match             9.2%;  Score 173;  DB 2;  Length 810;
  Best Local Similarity  24.8%;  Pred. No. 5.22e-12;
  Matches    31;  Conservative   41;  Mismatches  43;  Indels  10;  Gaps   9;

Db     377 CYHGDGQSYRGTSSTTTGKKCQSWSSMTPHWHEKTPENFPNAGLTMNYCRNPDAD-KGP 435
           |:  :|: ||   :: : | :| :|  : ::    ::   : ||   :|||||| | :||
Qy      25 CFWDNGHLYREDQTSPAPGLRCLNW--LDAQ-SGLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db     436 WCFTT-DPSVR-WEYCNLKKCSGTEGSVAAPPPVAQLPDA-ETPSEEDC-MFGNGKGYRG 491
           ||: : :::|     |:  :|: |  | | |:  ::: :| | |: ::  :|: :::  :
Qy      81 WCYVSGEAGVPEKRPCEDLRCPET-TSQALPAFTTEIQEASEGPGADEVQVFAPANALPA 139

Db     492 KKATT 496
           :  ::
Qy     140 RSEAA 144


RESULT   10
ENTRY            PLPG       #type fragment
TITLE            plasmin (EC 3.4.21.7) precursor - pig (fragment)
ALTERNATE_NAMES  plasminogen
CONTAINS         miniplasminogen
ORGANISM         #formal_name Sus scrofa domestica #common_name domestic pig
DATE             07-Sep-1990 #sequence_revision 01-Nov-1996 #text_change
                   18-Jul-1997
ACCESSIONS       S03733; S03737; A25834
REFERENCE        S03733
   #authors      Schaller, J.; Marti, T.; Roesselet, S.J.; Kaempfer, U.;
                   Rickli, E.E.
   #journal      Fibrinolysis (1987) 1:91-102
   #title        Amino acid sequence of the heavy chain of porcine plasmin.
                   Comparison of the carbohydrate attachment sites with the
                   human and bovine species.
   #accession    S03733
      ##molecule_type protein
      ##residues      1-560 ##label SCH
REFERENCE        S03735
   #authors      Brunisholz, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.;
                   Lergier, W.; Manneberg, M.; Gillessen, D.
   #journal      Eur. J. Biochem. (1981) 114:465-470
   #title        Comparison of the primary structure of the N-terminal CNBr
                   fragments of human, bovine and porcine plasminogen.
   #cross-references MUID:81212097
   #accession    S03737
      ##molecule_type protein
      ##residues      1-57 ##label BRU
REFERENCE        A25834
   #authors      Marti, T.; Schaller, J.; Rickli, E.E.
   #journal      Eur. J. Biochem. (1985) 149:279-285
   #title        Determination of the complete amino-acid sequence of porcine
                   miniplasminogen.
   #cross-references MUID:85203907
   #accession    A25834
      ##molecule_type protein
      ##residues      450-790 ##label MAR
FUNCTION
   #description  dissolves the fibrin of blood clots; acts as a proteolytic
                   factor in a variety of processes including embryonic
                   development, tissue remodeling and tumor invasion; in
                   ovulation it weakens the walls of the graafian follicle;
                   also activates the urokinase-type plasminogen activator
   #pathway      fibrinolysis
CLASSIFICATION   #superfamily plasmin; kringle homology; plasminogen-related
                   protein precursor homology; trypsin homology
KEYWORDS         fibrinolysis; glycoprotein; hydrolase; kidney; kringle;
                   plasma; serine proteinase
FEATURE
   1-790             #product plasminogen #status predicted #label PRO\
   1-77              #domain plasminogen-related protein precursor homology
                       (fragment) #label PLPH\
   1-77              #domain activation peptide #status predicted #label APT\
   78-560            #product plasmin chain A #status predicted #label ACH\
   84-162            #domain kringle homology #label KR1\
   166-243           #domain kringle homology #label KR2\
   256-333           #domain kringle homology #label KR3\
   358-435           #domain kringle homology #label KR4\
   450-790           #product miniplasminogen #status experimental #label
                       MIN\
   461-540           #domain kringle homology #label KR5\
   561-790           #product plasmin chain B #status experimental #label
                       BCH\
   561-783           #domain trypsin homology #label TRY\
   30-54,34-42,84-162,
   105-145,133-157,
   166-243,169-297,
   187-226,215-238,
   256-333,277-316,
   305-328,358-435,
   379-418,407-430,
   461-540,482-523,
   511-535,547-665,
   557-565,587-603,
   679-746,709-725,
   736-764           #disulfide_bonds #status predicted\
   602,645,740       #active_site His, Asp, Ser #status predicted
SUMMARY          #length 790  #checksum 3770

  Query Match             9.1%;  Score 171;  DB 1;  Length 790;
  Best Local Similarity  26.4%;  Pred. No. 1.13e-11;
  Matches    34;  Conservative   36;  Mismatches  49;  Indels  10;  Gaps   9;

Db     358 CYRGNGESYRGTSSTTITGRKCQSWVSMTPHRHEKTPGNFPNAGLTMNYCRNPDAD-KSP 416
           |:: ||  ||   ::   | :| :|  : ::    :::  : ||   :|||||| | ::|
Qy      25 CFWDNGHLYREDQTSPAPGLRCLNW--LDAQS-GLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db     417 WCYTT-DPRVR-WEYCNLKKCSET-EQQVTNFPAIAQVPSVEDLSEDC-MFGNGKRYRGK 472
           ||| : :: |     |:  :|:||  | :  | :  | :|    :::  :|: ::   ::
Qy      81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQVFAPANALPAR 140

Db     473 R-ATTVAGV 480
           |::|  |
Qy     141 SEAAAVQPV 149


RESULT   11
ENTRY            PLBO       #type complete
TITLE            plasmin (EC 3.4.21.7) precursor - bovine
ALTERNATE_NAMES  plasminogen
ORGANISM         #formal_name Bos primigenius taurus #common_name cattle
DATE             30-Sep-1987 #sequence_revision 28-Apr-1995 #text_change
                   05-Sep-1997
ACCESSIONS       S45046; A25835; I45961; S03736
REFERENCE        S45046
   #authors      Berglund, L.; Andersen, M.D.; Petersen, T.E.
   #submission   submitted to the EMBL Data Library, May 1994
   #description  Cloning and characterizatin of the bovine plasminogen cDNA.
   #accession    S45046
      ##molecule_type mRNA
      ##residues      1-812 ##label BER
      ##cross-references EMBL:X79402; NID:g494962; PID:g494963
      ##experimental_source liver
```

```
      ##note          it is uncertain whether Met-1 or Met-8 is the initiator
REFERENCE         A25835
   #authors       Schaller, J.; Moser, P.W.; Dannegger-Muller, G.A.K.;
                    Rosselet, S.J.; Kampfer, U.; Rickli, E.E.
   #journal       Eur. J. Biochem. (1985) 149:267-278
   #title         Complete amino acid sequence of bovine plasminogen.
                    Comparison with human plasminogen.
   #cross-references MUID:85203906
   #accession     A25835
      ##molecule_type protein
      ##residues      27-334,'D',336-515,'H',517-554,'L',556-812 ##label SCH
REFERENCE         I45961
   #authors       Malinowski, D.P.; Sadler, J.E.; Davie, E.W.
   #journal       Biochemistry (1984) 23:4243-4250
   #title         Characterization of a complementary deoxyribonucleic acid
                    coding for human and bovine plasminogen.
   #cross-references MUID:85023311
   #accession     I45961
      ##status        translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      706-743,'R',745-812 ##label MAL
      ##cross-references GB:K02935; NID:g163551; PID:g163552
REFERENCE         S03735
   #authors       Brunisholz, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.;
                    Lergier, W.; Manneberg, M.; Gillessen, D.
   #journal       Eur. J. Biochem. (1981) 114:465-470
   #title         Comparison of the primary structure of the N-terminal CNBr
                    fragments of human, bovine and porcine plasminogen.
   #cross-references MUID:81212097
   #accession     S03736
      ##molecule_type protein
      ##residues      27-83 ##label BRU
FUNCTION
   #description   dissolves the fibrin of blood clots; acts as a proteolytic
                    factor in a variety of processes including embryonic
                    development, tissue remodeling and tumor invasion; in
                    ovulation it weakens the walls of the graafian follicle;
                    also activates the urokinase-type plasminogen activator
   #pathway       fibrinolysis
CLASSIFICATION    #superfamily plasmin; kringle homology; plasminogen-related
                    protein precursor homology; trypsin homology
KEYWORDS          duplication; fibrinolysis; glycoprotein; hydrolase; kidney;
                    kringle; plasma; serine proteinase; zymogen
FEATURE
   1-26               #domain signal sequence #status predicted #label SIG\
   8-103              #domain plasminogen-related protein precursor homology
                        #label PLPH\
   27-812             #product plasminogen #status experimental #label PRO\
   27-103             #domain activation peptide #status experimental #label
                        APT\
   104-583,584-812    #product plasmin #status experimental #label MAT\
   104-583            #domain plasmin chain A #status experimental #label ACH\
   110-188            #domain kringle homology #label KR1\
   192-269            #domain kringle homology #label KR2\
   282-359            #domain kringle homology #label KR3\
   384-461            #domain kringle homology #label KR4\
   485-564            #domain kringle homology #label KR5\
   584-812            #domain plasmin chain B #status experimental #label BCH\
   584-805            #domain trypsin homology #label TRY\
   56-80,60-68,
   110-188,131-171,
   159-183,192-269,
   195-323,213-252,
   241-264,282-359,
   303-342,331-354,
   384-461,405-444,
   433-456,485-564,
   506-547,535-559,
   570-687,580-588,
   609-625,701-768,
   731-747,758-786    #disulfide_bonds #status predicted\
   315                #binding_site carbohydrate (Asn) (covalent) #status
                        experimental\
   365                #binding_site carbohydrate (Ser) (covalent) #status
                        experimental\
   624,667,762        #active_site His, Asp, Ser #status predicted
SUMMARY           #length 812  #molecular-weight 91215  #checksum 321

  Query Match             9.1%;  Score 172;  DB 1;  Length 812;
  Best Local Similarity  31.1%;  Pred. No. 7.67e-12;
  Matches    33;  Conservative   25;  Mismatches  39;  Indels   9;  Gaps   8;

Db     384 CYHGNGQSYRGTSSTTITGRKCQSWSSMTPHRHLKTPENYPNAGLTMNYCRNPDAD-KSP 442
           |:  ||: ||   ::   | :| :|  : ::  | ::   : ||   :|||||| | ::|
Qy      25 CFWDNGHLYREDQTSPAPGLRCLNW--LDAQSGL-ASAPVSGAG-NHSYCRNPDEDPRGP 80

Db     443 WCYTT-DPRVR-WEFCNLKKCSETPEQ-VPA-APQAPGVENPPEAD 484
           ||| : :: |     |:  :|:||  | :|| : :     : | ||
Qy      81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGAD 126


RESULT   12
ENTRY             A60185     #type complete
TITLE             hepatocyte growth factor precursor - mouse
ALTERNATE_NAMES   hepapoietin A; scatter factor
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              03-Mar-1993 #sequence_revision 26-May-1994 #text_change
                    26-Feb-1999
ACCESSIONS        JC2117; PC2064; A60185; S43416; S45521; S17173; S10966;
                    I48758; JU0231
REFERENCE         JC2117
   #authors       Sasaki, M.; Nishio, M.; Sasaki, T.; Enami, J.
   #journal       Biochem. Biophys. Res. Commun. (1994) 199:772-779
   #title         Identification of mouse mammary fibroblast-derived mammary
                    growth factor as hepatocyte growth factor.
   #cross-references MUID:94183257
   #accession     JC2117
      ##molecule_type mRNA
      ##residues      1-728 ##label SAS2
      ##cross-references GB:D10212; NID:g220435; PID:d1001533; PID:g220436
      ##experimental_source fibroblast, COS-1 cell
      ##note          submitted to JIPID, May 1993
   #accession     PC2064
      ##molecule_type protein
      ##residues      496-504 ##label SA2
REFERENCE         A60185
   #authors       Rosen, E.M.; Meromsky, L.; Setter, E.; Vinter, D.W.;
                    Goldberg, I.D.
   #journal       Proc. Soc. Exp. Biol. Med. (1990) 195:34-43
   #title         Purified scatter factor stimulates epithelial and vascular
                    endothelial cell migration (43115).
   #cross-references MUID:90377927
   #accession     A60185
      ##molecule_type protein
      ##residues      'X',184-188,'KX',191-192,'X',194,'XX',197;357-364,'XX',
                       367;375-377,'E',379,'E',381-383;653-666 ##label ROS
REFERENCE         S43416
   #authors       Liu, Y.; Michalopoulos, G.K.; Zarnegar, R.
   #journal       Biochim. Biophys. Acta (1993) 1216:299-303
   #title         Molecular cloning and characterization of cDNA encoding mouse
                    hepatocyte growth factor.
   #cross-references MUID:94060105
   #accession     S43416
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-728 ##label LIU
      ##cross-references EMBL:X72307
REFERENCE         S45521
   #authors       Liu, Y.
   #submission    submitted to the EMBL Data Library, May 1993
   #accession     S45521
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-563,'H',565-728 ##label LI2
      ##cross-references EMBL:X72307
REFERENCE         S17173
```

```
   #authors       Coffer, A.; Fellows, J.; Young, S.; Pappin, D.; Rahman, D.
   #journal       Biochem. J. (1991) 278:35-41
   #title         Purification and characterization of biologically active
                    scatter factor from ras-transformed NIH 3T3 conditioned
                    medium.
   #cross-references MUID:91354223
   #accession     S17173
      ##molecule_type protein
      ##residues      496-517,'T',519 ##label COF
REFERENCE         S10966
   #authors       Gherardi, E.; Stoker, M.
   #journal       Nature (1990) 346:228
   #title         Hepatocytes and scatter factor.
   #cross-references MUID:90326152
   #accession     S10966
      ##status        preliminary
      ##molecule_type protein
      ##residues      496-507,'X',509-512,'L',514-516,'X',518-519 ##label NAT
REFERENCE         I48758
   #authors       Plaschke-Schlutter, A.; Behrens, J.; Gherardi, E.;
                    Birchmeier, W.
   #journal       J. Biol. Chem. (1995) 270:830-836
   #title         Characterization of the scatter factor/hepatocyte growth
                    factor gene promoter. Positive and negative regulatory
                    elements direct gene expression to mesenchymal cells.
   #cross-references MUID:95122532
   #accession     I48758
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues      1-30 ##label RES
      ##cross-references EMBL:X81630; NID:g673451; PID:g673452
COMPLEX           disulfide-bonded heterodimer of chains derived from the same
                    precursor
FUNCTION
   #description   stimulates mitosis of hepatocytes and other cells
   #note          does not have proteinase activity
CLASSIFICATION    #superfamily hepatocyte growth factor; kringle homology;
                    trypsin homology
KEYWORDS          alternative splicing; glycoprotein; growth factor;
                    heterodimer; kringle; pyroglutamic acid

FEATURE
   1-32               #domain signal sequence #status predicted #label SIG\
   56-495,496-728     #product hepatocyte growth factor #status predicted
                        #label MAT\
   56-495             #domain hepatocyte growth factor alpha chain #status
                        predicted #label ACH\
   129-207            #domain kringle homology #label KR1\
   212-289            #domain kringle homology #label KR2\
   306-384            #domain kringle homology #label KR3\
   392-470            #domain kringle homology #label KR4\
   496-728            #domain hepatocyte growth factor beta chain #status
                        predicted #label BCH\
   496-719            #domain trypsin homology #label TRY\
   33                 #modified_site pyrrolidone carboxylic acid (Gln) (in
                        mature form) #status predicted\
   295,403,569,656    #binding_site carbohydrate (Asn) (covalent) #status
                        predicted\
   488-607            #disulfide_bonds #status predicted
SUMMARY           #length 728  #molecular-weight 82944  #checksum 4034

  Query Match             8.9%;  Score 168;  DB 1;  Length 728;
  Best Local Similarity  31.9%;  Pred. No. 3.54e-11;
  Matches    30;  Conservative   23;  Mismatches  34;  Indels   7;  Gaps   5;

Db     387 SSGQDCYRGNGKNYMGNLSKTRSGLTCSMWDKNMEDLHRHIFWEPDASKLNKNYCRNPDD 446
           :||  |:: ||  |  : :   :|| |  |   ::  :  :   | ::  | :||||||:
Qy      21 GSG-GCFWDNGHLYREDQTSPAPGLRCLNW---LDA-QSGLASAPVSGAGNHSYCRNPDE 75

Db     447 DAHGPWCY-TGNPLIP-WDYCPISRCEGDTTPTI 478
           |::|||| :|:: :|    |   ||   |: ::
Qy      76 DPRGPWCYVSGEAGVPEKRPCEDLRCPETTSQAL 109
```

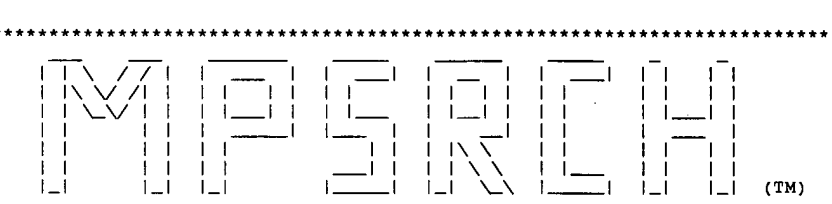
```
RESULT   13
ENTRY             JH0579     #type complete
TITLE             hepatocyte growth factor precursor - human
ALTERNATE_NAMES   hepapoietin A; scatter factor
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change
                    18-Sep-1998
ACCESSIONS        JH0579; JU0333; A41140; B36677; A36677; A33512; A39006;
                    PH0114; A37796; S06794; I59214; S15443
REFERENCE         JH0579
   #authors       Seki, T.; Hagiya, M.; Shimonishi, M.; Nakamura, T.; Shimizu,
                    S.
   #journal       Gene (1991) 102:213-219
   #title         Organization of the human hepatocyte growth factor-encoding
                    gene.
   #cross-references MUID:91340155
   #accession     JH0579
      ##molecule_type DNA
      ##residues      1-728 ##label SEK
      ##cross-references DDBJ:D90318
      ##note          the authors translated the codon GAA for residue 662 as
                       Gly
REFERENCE         JU0333
   #authors       Seki, T.; Hagiya, M.; Shimonishi, M.; Nakamura, T.; Shimizu,
                    S.
   #submission    submitted to JIPID, March 1991
   #description   Organization of the human hepatocyte growth factor-encoding
                    gene.
   #accession     JU0333
      ##molecule_type DNA
      ##residues      1-481,'RT',484-728 ##label SE2
REFERENCE         A41140
   #authors       Weidner, K.M.; Arakaki, N.; Hartmann, G.; Vandekerckhove, J.;
                    Weingart, S.; Rieder, H.; Fonatsch, C.; Tsubouchi, H.;
                    Hishida, T.; Daikuhara, Y.; Birchmeier, W.
   #journal       Proc. Natl. Acad. Sci. U.S.A. (1991) 88:7001-7005
   #title         Evidence for the identity of human scatter factor and human
                    hepatocyte growth factor.
   #cross-references MUID:91334393
   #accession     A41140
      ##molecule_type mRNA
      ##residues      1-728 ##label WEI
      ##cross-references GB:M73239; NID:g337935; PID:g337936
REFERENCE         A36677
   #authors       Seki, T.; Ihara, I.; Sugimura, A.; Shimonishi, M.; Nishizawa,
                    T.; Asami, O.; Hagiya, M.; Nakamura, T.; Shimizu, S.
   #journal       Biochem. Biophys. Res. Commun. (1990) 172:321-327
   #title         Isolation and expression of cDNA for different forms of
                    hepatocyte growth factor from human leukocyte.
   #cross-references MUID:91025062
   #accession     B36677
      ##molecule_type mRNA
      ##residues      1-728 ##label SE3
      ##cross-references GB:M60718; NID:g184031; PID:g184032
   #accession     A36677
      ##molecule_type mRNA
      ##residues      1-161,167-728 ##label SE4
      ##cross-references EMBL:X16323
      ##experimental_source leukocyte
REFERENCE         A33512
   #authors       Miyazawa, K.; Tsubouchi, H.; Naka, D.; Takahashi, K.;
                    Okigaki, M.; Arakaki, N.; Nakayama, H.; Hirono, S.;
                    Sakiyama, O.; Takahashi, K.; Gohda, E.; Daikuhara, Y.;
                    Kitamura, N.
   #journal       Biochem. Biophys. Res. Commun. (1989) 163:967-973
   #title         Molecular cloning and sequence analysis of cDNA for human
                    hepatocyte growth factor.
   #cross-references MUID:89392017
   #accession     A33512
      ##status        not compared with conceptual translation
      ##molecule_type mRNA
      ##residues      1-728 ##label MIY
      ##cross-references GB:M29145; NID:g184041; PID:g306846
```

```
REFERENCE        A39006
   #authors      Rubin, J.S.; Chan, A.M.L.; Bottaro, D.P.; Burgess, W.H.;
                   Taylor, W.G.; Cech, A.C.; Hirschfield, D.W.; Wong, J.;
                   Miki, T.; Finch, P.W.; Aaronson, S.A.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1991) 88:415-419
   #title        A broad-spectrum human lung fibroblast-derived mitogen is a
                   variant of hepatocyte growth factor.
   #cross-references MUID:91110540
   #accession    A39006
      ##molecule_type mRNA
      ##residues      1-161,167-728 ##label RUB
      ##cross-references GB:M55379
      ##experimental_source embryonic lung
REFERENCE        PH0114
   #authors      Yoshiyama, Y.; Arakaki, N.; Naka, D.; Takahashi, K.; Hirono,
                   S.; Kondo, J.; Nakayama, H.; Gohda, E.; Kitamura, N.;
                   Tsubouchi, H.; Ishii, T.; Hishida, T.; Daikuhara, Y.
   #journal      Biochem. Biophys. Res. Commun. (1991) 175:660-667
   #title        Identification of the N-terminal residue of the heavy chain
                   of both native and recombinant human hepatocyte growth
                   factor.
   #cross-references MUID:91207365
   #accession    PH0114
      ##molecule_type protein
      ##residues      32-43;53-58 ##label YOS
      ##experimental_source plasma
REFERENCE        A37796
   #authors      Weidner, K.M.; Behrens, J.; Vandekerckhove, J.; Birchmeier,
                   W.
   #journal      J. Cell Biol. (1990) 111:2097-2108
   #title        Scatter factor: molecular characteristics and effect on the
                   invasiveness of epithelial cells.
   #cross-references MUID:91035621
   #accession    A37796
      ##molecule_type protein
      ##residues      86-91;329-344;356-363,'XX',366-370;425-434;442-447,'X',
                        449-450;543-546,'X',548-553;563-565,'X',567-574
                        ##label WE2
REFERENCE        S06794
   #authors      Nakamura, T.; Nishizawa, T.; Hagiya, M.; Seki, T.;
                   Shimonishi, M.; Sugimura, A.; Tashiro, K.; Shimizu, S.
   #journal      Nature (1989) 342:440-443
   #title        Molecular cloning and expression of human hepatocyte growth
                   factor.
   #cross-references MUID:90066676
   #accession    S06794
      ##molecule_type mRNA
      ##residues      1-31,'HK',34-77,'N',79-292,'V',294-299,'M',301-316,'A',
                        318-335,'K',337-386,'N',388-415,'N',417-504,'V',
                        506-508,'I',510-557,'E',559-560,'R',562-594,'N',
                        596-728 ##label NAK
      ##cross-references EMBL:X16323; NID:g32081; PID:g32082
      ##experimental_source liver
      ##note          the authors translated the codon CAG for residue 727 as
                        Glu
      ##note          part of this sequence, including the amino end of both
                        the alpha and beta chains, was confirmed by protein
                        sequencing
REFERENCE        I59214
   #authors      Hartmann, G.; Naldini, L.; Weidner, K.M.; Sachs, M.; Vigna,
                   E.; Comoglio, P.M.; Birchmeier, W.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1992) 89:11574-11578
   #title        A functional domain in the heavy chain of scatter
                   factor/hepatocyte growth factor binds the c-Met receptor
                   and induces cell dissociation but not mitogenesis.
   #cross-references MUID:93087571
   #accession    I59214
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      1-288,'ET' ##label RES
      ##cross-references GB:L02931; NID:g184033; PID:g184034
REFERENCE        S15443
   #authors      Miyazawa, K.; Kitamura, A.; Naka, D.; Kitamura, N.
   #journal      Eur. J. Biochem. (1991) 197:15-22
   #title        An alternatively processed mRNA generated from human
                   hepatocyte growth factor gene.
   #cross-references MUID:91200041
   #accession    S15443
      ##status       preliminary
      ##molecule_type mRNA
      ##residues      1-288,'ET' ##label MI2
      ##cross-references EMBL:X57574; NID:g32083; PID:g32084
GENETICS
   #gene         GDB:HGF
      ##cross-references GDB:127524; OMIM:142409
   #map_position 7q21.1-7q21.1
   #introns      30/1; 85/2; 123/1; 161/2; 209/1; 249/2; 289/1; 347/2; 390/1;
                   424/2; 469/1; 482/1; 514/2; 539/2; 586/2; 622/1; 670/3
COMPLEX          disulfide-bonded heterodimer of chains derived from the same
                   precursor
FUNCTION
   #description  stimulates mitosis of hepatocytes and other cells
   #note         does not have proteinase activity
CLASSIFICATION   #superfamily hepatocyte growth factor; kringle homology;
                   trypsin homology
KEYWORDS         alternative splicing; glycoprotein; growth factor;
                   heterodimer; kringle; pyroglutamic acid
FEATURE
   1-31              #domain signal sequence #status predicted #label SIG\
   32-494,495-728    #product hepatocyte growth factor #status experimental
                       #label MAT\
   32-494            #domain alpha chain #status experimental #label ACH\
   128-206           #domain kringle homology #label KR1\
   211-288           #domain kringle homology #label KR2\
   305-383           #domain kringle homology #label KR3\
   391-469           #domain kringle homology #label KR4\
   495-728           #domain beta chain #status experimental #label BCH\
   495-716           #domain trypsin homology #label TRY\
   32                #modified_site pyrrolidone carboxylic acid (Gln) (in
                       mature form) #status experimental\
   294,402,566,653   #binding_site carbohydrate (Asn) (covalent) #status
                       predicted\
   487-604           #disulfide_bonds #status predicted
SUMMARY          #length 728  #molecular-weight 83133  #checksum 4163

  Query Match             8.9%;  Score 167;  DB 1;  Length 728;
  Best Local Similarity  31.5%;  Pred. No. 5.19e-11;
  Matches    28;  Conservative   22;  Mismatches  33;  Indels   6;  Gaps   4;

Db    391 CYRGNGKNYMGNLSQTRSGLTCSMWDKNMEDLHRHIFWEPDASKLNENYCRNPDDDAHGP 450
          |:: || |  : :   :|| |  |   ::  :  :   | ::  | :||||||:|::||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW---LDA-QSGLASAPVSGAGNHSYCRNPDEDPRGP 80

Db    451 WCY-TGNPLIP-WDYCPISRCEGDTTPTI 477
          ||| :|:: :|    |   ||   |: ::
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQAL 109


RESULT   14
ENTRY            TBHU       #type complete
TITLE            thrombin (EC 3.4.21.5) precursor - human
ALTERNATE_NAMES  coagulation factor II
CONTAINS         prothrombin
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             30-Nov-1980 #sequence_revision 22-Jul-1994 #text_change
                   20-Mar-1998
ACCESSIONS       A29351; A00914; B00914; A37549; A37550; I51952
REFERENCE        A29351
   #authors      Degen, S.J.F.; Davie, E.W.
   #journal      Biochemistry (1987) 26:6165-6177
   #title        Nucleotide sequence of the gene for human prothrombin.
   #cross-references MUID:88077877
   #accession    A29351
      ##molecule_type DNA
      ##residues      1-622 ##label DEG
      ##cross-references GB:M17262; GB:M33691; NID:g558069; PID:g339641
```

```
REFERENCE        A00914
   #authors      Degen, S.J.F.; MacGillivray, R.T.A.; Davie, E.W.
   #journal      Biochemistry (1983) 22:2087-2097
   #title        Characterization of the complementary deoxyribonucleic acid
                   and gene coding for human prothrombin.
   #cross-references MUID:83231469
   #accession    A00914
      ##molecule_type mRNA
      ##residues      8-163,'N',165-622 ##label DE2
      ##cross-references GB:V00595; GB:J00307; NID:g37128; PID:e5121;
                        PID:g1335344
   #accession    B00914
      ##molecule_type DNA
      ##residues      188-311 ##label DE3
REFERENCE        A37549
   #authors      Walz, D.A.; Hewett-Emmett, D.; Seegers, W.H.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1977) 74:1969-1972
   #cross-references MUID:77193964
   #accession    A37549
      ##molecule_type protein
      ##residues      44-118,'N',120,'S',122-163,'I',165-175,'A',177-182,'T',
                       184-193,'MV',196-308,'EE',309-314 ##label WAL
REFERENCE        A37550
   #authors      Butkowski, R.J.; Elion, J.; Downing, M.R.; Mann, K.G.
   #journal      J. Biol. Chem. (1977) 252:4942-4957
   #title        Primary structure of human prethrombin 2 and alpha-thrombin.
   #cross-references MUID:77207112
   #accession    A37550
      ##molecule_type protein
      ##residues      315-334,'N',336-348,'N',350-368,'N',370-397,'N',399-413,
                       'N',415-484,'N',486-493,'G',495-503,'Y',505-508,'S',
                       510,'V',512-513,'D',515-528,'AL',531,'Q',533-622
                       ##label BUT
REFERENCE        A37551
   #authors      Rabiet, M.J.; Blashill, A.; Furie, B.; Furie, B.C.
   #journal      J. Biol. Chem. (1986) 261:13210-13215
   #cross-references MUID:87008532
   #contents     annotation; activation cleavages
REFERENCE        I51952
   #authors      MacGillivray, R.T.; Irwin, D.M.; Guinto, E.R.; Stone, J.C.
   #journal      Ann. N. Y. Acad. Sci. (1986) 485:73-79
   #title        Recombinant genetic approaches to functional mapping of
                  thrombin.
   #cross-references MUID:87182874
   #accession    I51952
      ##status       translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      1-2,'RI',5-100 ##label RES
      ##cross-references GB:M33031; NID:g190723; PID:g190724
COMMENT     Thrombin, which cleaves bonds after Arg and Lys, converts
              fibrinogen to fibrin and activates factors V, VIII, XIII, and, in
              complex with thrombomodulin, protein C.
COMMENT     Prothrombin is activated on the surface of a phospholipid membrane
              that binds the amino end of prothrombin and factors Va and Xa in
              calcium-dependent interactions. The activation peptide(s) can be
              removed either by factor Xa or thrombin; the cleavage into light
              and heavy chains is by factor Xa. It is not known whether one or
              two smaller activation peptides, with additional cleavage after
              314-Arg, are released in natural blood clotting.
COMMENT     The cleavage after Arg-198, observed in vitro, does not occur in
              plasma.
COMMENT     The gamma-carboxyglutamyl residues bind calcium ions, result from
              the carboxylation of glutamyl residues by microsomal vitamin
              K-dependent carboxylase, and are necessary for calcium-dependent
              interaction with the negatively charged phospholipid membrane
              surface.
COMMENT     The prothrombin precursor is synthesized in the liver.
GENETICS
   #gene         GDB:F2
      ##cross-references GDB:119894; OMIM:176930
   #map_position 11p11-11q12
   #introns      27/1; 80/3; 89/1; 106/1; 141/2; 187/1; 292/1; 335/1; 377/2;
                   433/2; 491/2; 552/1; 575/3
CLASSIFICATION   #superfamily thrombin; Gla domain homology; kringle homology;
                   trypsin homology
KEYWORDS         acute phase; blood coagulation; calcium binding;
                   carboxyglutamic acid; duplication; glycoprotein; hydrolase;
                   kringle; liver; plasma; serine proteinase
FEATURE
   1-24              #domain signal sequence #status predicted #label SIG\
   25-43             #domain propeptide #status predicted #label PRO\
   28-87             #domain Gla domain homology #label GLA\
   44-622            #product prothrombin #status experimental #label MAT\
   44-327            #domain activation peptide #status experimental #label
                       APT\
   108-186           #domain kringle homology #label KR1\
   213-291           #domain kringle homology #label KR2\
   328-363           #product thrombin light chain #status experimental
                       #label LCH\
   364-622           #product thrombin heavy chain #status experimental
                       #label HCH\
   364-613           #domain trypsin homology #label TRY\
   49,50,57,59,62,63,
   68,69,72,75       #modified_site gamma-carboxyglutamic acid (Glu) #status
                       experimental\
   60-65,90-103,
   108-186,129-169,
   157-181,213-291,
   234-274,262-286   #disulfide_bonds #status predicted\
   121,143           #binding_site carbohydrate (Asn) (covalent) #status
                       predicted\
   336-482,536-550,
   564-594           #disulfide_bonds #status predicted\
   391-407           #disulfide_bonds #status experimental\
   406,462           #active_site His, Asp #status predicted\
   416               #binding_site carbohydrate (Asn) (covalent) #status
                       experimental\
   568               #active_site Ser #status experimental
SUMMARY          #length 622  #molecular-weight 70036  #checksum 3003

  Query Match             8.8%;  Score 165;  DB 1;  Length 622;
  Best Local Similarity  37.7%;  Pred. No. 1.11e-10;
  Matches    26;  Conservative   13;  Mismatches  26;  Indels   4;  Gaps   3;

Db    213 CVPDRGQQYQGRLAVTTHGLPCLAWASAQAKALSKHQDFNSAVQLVENFCRNPDGDEEGV 272
          |  | |: |:   :  : || || |  ||: :|:     ::|     ::|||||  |  |
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLDAQS-GLASAP-VSGAGN--HSYCRNPDEDPRGP 80

Db    273 WCYVAGKPG 281
          ||||:| :|
Qy     81 WCYVSGEAG 89


RESULT   15
ENTRY            A47136     #type complete
TITLE            macrophage-stimulating protein 1 precursor - human
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             03-May-1994 #sequence_revision 14-Nov-1997 #text_change
                   11-Sep-1998
ACCESSIONS       A40331; B40331; A47136; A61395
REFERENCE        A40331
   #authors      Han, S.; Stuart, L.A.; Degen, S.J.F.
   #journal      Biochemistry (1991) 30:9768-9780
   #title        Characterization of the DNF15S2 locus on human chromosome 3:
                   identification of a gene coding for four kringle domains
                   with homology to hepatocyte growth factor.
   #cross-references MUID:92002016
   #accession    A40331
      ##molecule_type DNA
      ##residues      1-711 ##label HA1
      ##cross-references GB:M74179
   #accession    B40331
      ##molecule_type mRNA
      ##residues      1-711 ##label HA2
      ##cross-references GB:M74178; NID:g183976; PID:g183977
REFERENCE        A47136
```

```
   #authors       Yoshimura, T.; Yuhki, N.; Wang, M.H.; Skeel, A.; Leonard,
                    E.J.
   #journal       J. Biol. Chem. (1993) 268:15461-15468
   #title         Cloning, sequencing, and expression of human macrophage
                    stimulating protein (MSP, MST1) confirms MSP as a member of
                    the family of kringle proteins and locates the MSP gene on
                    chromosome 3.
   #cross-references MUID:93340141
   #accession     A47136
      ##molecule_type mRNA
      ##residues      1-12,'C',14-622,'F',624-711 ##label YOS
      ##cross-references GB:L11924; NID:g398037; PID:g398038
      ##note          authors translated the codon TTT for residue 623 as Leu;
                       parts of this sequence were determined by protein
                       sequencing
REFERENCE         A61395
   #authors       Skeel, A.; Yoshimura, T.; Showalter, S.D.; Tanaka, S.;
                    Appella, E.; Leonard, E.J.
   #journal       J. Exp. Med. (1991) 173:1227-1234
   #title         Macrophage stimulating protein: purification, partial amino
                    acid sequence, and cellular activity.
   #cross-references MUID:91217635
   #accession     A61395
      ##molecule_type protein
      ##residues      230-247;288-291,'E',293-295,'X',297-301,'X',303,'E',305,
                       'EX',308-310;326-331,'X',333-341;484-501;530-536,'X',
                       538-546,'X',548-549,'SL';574-592,'E',594-596;602-611
                       ##label SKE
      ##experimental_source plasma
GENETICS
   #gene          GDB:MST1; D3F15S2; DNF15S2; HGFL
     .##cross-references GDB:128833; OMIM:142408
   #map_position 3p21-3p21.3
COMPLEX           disulfide-bonded heterodimer of chains derived from the same
                    precursor
CLASSIFICATION    #superfamily hepatocyte growth factor; kringle homology;
                    trypsin homology
KEYWORDS          duplication; glycoprotein; growth factor; kringle; plasma
FEATURE
   1-18               #domain signal sequence #status predicted #label SIG\
   19-483,484-711     #product macrophage-stimulating protein 1 #status
                        predicted #label MAT\
   19-483             #domain alpha chain #status predicted #label ACH\
   110-186            #domain kringle homology #label KR1\
   191-268            #domain kringle homology #label KR2\
   283-361            #domain kringle homology #label KR3\
   370-448            #domain kringle homology #label KR4\
   484-711            #domain beta chain #status predicted #label BCH\
   484-704            #domain trypsin homology #label TRY\
   56-78,60-66,
   110-186,131-169,
   157-181,191-268,
   212-251,240-263,
   283-361,304-343,
   332-355,370-448,
   391-431,419-443,
   507-523,602-667,
   632-646,657-685    #disulfide_bonds #status predicted\
   72,296,615         #binding_site carbohydrate (Asn) (covalent) #status
                        predicted
SUMMARY           #length 711  #molecular-weight 80379  #checksum 5521

  Query Match             8.8%;  Score 166;  DB 1;  Length 711;
  Best Local Similarity  32.5%;  Pred. No. 7.58e-11;
  Matches    27;  Conservative   17;  Mismatches  35;  Indels   4;  Gaps   4;

Db    110 CIMNNGVGYRGTMATTVGGLPCQAWSHKFPNDHKYTPTLRNGLEENFCRNPDGDPGGPWC 169
          |: :||  ||   ::   || |  |     :    :|    | : ::||||| || ||||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLDA-QSGLASAPVSGAG-NHSYCRNPDEDPRGPWC 82

Db    170 YTT-DPAVRF-QSCGIKSCREAA 190
          | : :::|   ::|    | |::
Qy     83 YVSGEAGVPEKRPCEDLRCPETT 105
```

```
Search completed: Fri Sep 17 18:26:39 1999
Job time : 50 secs.
```



MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 17 18:26:56 1999; MasPar time 8.89 Seconds
835.926 Million cell updates/sec
Tabular output not generated.

Title: >US-09-084-491A-2
Description: (1-263) from US09084491A.pep
Perfect Score: 1883
Sequence: 1 MLLAWVQAFLVSNMLLAEAY.....PVDPQEGSTPLMGQAGTPGA 263

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 46.647; Variance 77.329; scale 0.603

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	204	10.8	169	1	PLMN_RAT	PLASMINOGEN (EC 3.4.21	1.10e-20
2	184	9.8	812	1	PLMN_MOUSE	PLASMINOGEN PRECURSOR	1.20e-16
3	183	9.7	4548	1	APOA_HUMAN	APOLIPOPROTEIN(A) PREC	1.89e-16
4	181	9.6	655	1	HGFA_HUMAN	HEPATOCYTE GROWTH FACT	4.70e-16
5	177	9.4	716	1	HGFL_MOUSE	HEPATOCYTE GROWTH FACT	2.88e-15
6	176	9.3	810	1	PLMN_ERIEU	PLASMINOGEN PRECURSOR	4.53e-15
7	173	9.2	810	1	PLMN_MACMU	PLASMINOGEN PRECURSOR	1.75e-14
8	171	9.1	790	1	PLMN_PIG	PLASMINOGEN (EC 3.4.21	4.27e-14
9	172	9.1	812	1	PLMN_BOVIN	PLASMINOGEN PRECURSOR	2.73e-14
10	168	8.9	728	1	HGF_MOUSE	HEPATOCYTE GROWTH FACT	1.62e-13
11	167	8.9	728	1	HGF_HUMAN	HEPATOCYTE GROWTH FACT	2.53e-13
12	165	8.8	622	1	THRB_HUMAN	PROTHROMBIN PRECURSOR	6.12e-13
13	166	8.8	711	1	HGFL_HUMAN	HEPATOCYTE GROWTH FACT	3.94e-13
14	165	8.8	728	1	HGF_RAT	HEPATOCYTE GROWTH FACT	6.12e-13
15	164	8.7	625	1	THRB_BOVIN	PROTHROMBIN PRECURSOR	9.51e-13
16	163	8.7	1420	1	APOA_MACMU	APOLIPOPROTEIN(A) (EC	1.48e-12
17	161	8.6	810	1	PLMN_HUMAN	PLASMINOGEN PRECURSOR	3.54e-12
18	158	8.4	603	1	FA12_CAVPO	COAGULATION FACTOR XII	1.31e-11
19	156	8.3	431	1	URTB_DESRO	SALIVARY PLASMINOGEN A	3.11e-11
20	156	8.3	477	1	URT2_DESRO	SALIVARY PLASMINOGEN A	3.11e-11
21	155	8.2	566	1	UROT_BOVIN	TISSUE PLASMINOGEN ACT	4.79e-11
22	152	8.1	333	1	PLMN_CANFA	PLASMINOGEN (EC 3.4.21	1.73e-10
23	152	8.1	433	1	UROK_PAPCY	UROKINASE-TYPE PLASMIN	1.73e-10
24	152	8.1	434	1	UROK_CHICK	UROKINASE-TYPE PLASMIN	1.73e-10
25	151	8.0	432	1	UROK_RAT	UROKINASE-TYPE PLASMIN	2.66e-10
26	150	8.0	442	1	UROK_PIG	UROKINASE-TYPE PLASMIN	4.07e-10
27	150	8.0	562	1	UROT_HUMAN	TISSUE PLASMINOGEN ACT	4.07e-10
28	149	7.9	593	1	FA12_BOVIN	COAGULATION FACTOR XII	6.22e-10
29	147	7.8	343	1	PLMN_SHEEP	PLASMINOGEN (EC 3.4.21	1.45e-09
30	147	7.8	394	1	URTG_DESRO	SALIVARY PLASMINOGEN A	1.45e-09
31	146	7.8	431	1	UROK_HUMAN	UROKINASE-TYPE PLASMIN	2.21e-09
32	147	7.8	433	1	UROK_MOUSE	UROKINASE-TYPE PLASMIN	1.45e-09
33	146	7.8	617	1	THRB_RAT	PROTHROMBIN PRECURSOR	2.21e-09
34	144	7.6	618	1	THRB_MOUSE	PROTHROMBIN PRECURSOR	5.10e-09
35	141	7.5	433	1	UROK_BOVIN	UROKINASE-TYPE PLASMIN	1.78e-08
36	141	7.5	559	1	UROT_MOUSE	TISSUE PLASMINOGEN ACT	1.78e-08
37	137	7.3	615	1	FA12_HUMAN	COAGULATION FACTOR XII	9.20e-08
38	135	7.2	325	1	PLMN_PETMA	PLASMINOGEN (EC 3.4.21	2.08e-07
39	134	7.1	477	1	URT1_DESRO	SALIVARY PLASMINOGEN A	3.11e-07
40	131	7.0	559	1	UROT_RAT	TISSUE PLASMINOGEN ACT	1.04e-06
41	126	6.7	338	1	PLMN_HORSE	PLASMINOGEN (EC 3.4.21	7.53e-06
42	105	5.6	451	1	PSS_ECOLI	CDP-DIACYLGLYCEROL--SE	1.89e-02
43	100	5.3	678	1	GSPD_AERHY	GENERAL SECRETION PATH	1.06e-01
44	99	5.3	678	1	GSPD_AERSA	GENERAL SECRETION PATH	1.49e-01
45	100	5.3	854	1	LDLR_CRIGR	LOW-DENSITY LIPOPROTEI	1.06e-01

ALIGNMENTS

RESULT 1
ID PLMN_RAT STANDARD; PRT; 169 AA.
AC Q01177;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE PLASMINOGEN (EC 3.4.21.7) (FRAGMENT).
GN PLG.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 91250378.
RA KANALAS J.J., MAKKER S.P.;
RT "Identification of the rat Heymann nephritis autoantigen (GP330) as a
RT receptor site for plasminogen.";
RL J. BIOL. CHEM. 266:10825-10829(1991).
CC -!- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
CC A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
CC EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION,
CC AND INFLAMMATION; IN OVULATION IT WEAKENS THE WALLS OF THE
CC GRAAFIAN FOLLICLE. IT ACTIVATES THE UROKINASE-TYPE PLASMINOGEN
CC ACTIVATOR, COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH
CC AS C1 AND C5. IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
CC LAMININ AND VON WILLEBRAND FACTOR.
CC -!- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN
CC ACTIVATORS, BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
CC FIBRIN. CANNOT BE ACTIVATED WITH STREPTOKINASE.
CC -!- PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN IMMEDIATELY AFTER
CC DISSOCIATION FROM THE CLOT.
CC -!- SIMILARITY: CONTAINS 5 KRINGLE REGIONS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
CC ----------

CC ----------
DR EMBL; M62832; G554488; -.
DR PIR; A40522; A40522.
DR PROSITE; PS00134; TRYPSIN_HIS; PARTIAL.

```
DR   PROSITE; PS00135; TRYPSIN_SER; PARTIAL.
DR   PROSITE; PS00021; KRINGLE_1; 1.
DR   PROSITE; PS50070; KRINGLE_2; 2.
DR   PFAM; PF00051; kringle; 1.
DR   HSSP; P00747; 1PMK.
KW   HYDROLASE; SERINE PROTEASE; PLASMA; GLYCOPROTEIN; FIBRINOLYSIS;
KW   TISSUE REMODELING; BLOOD COAGULATION; KRINGLE.
FT   NON_TER       1      1
FT   DOMAIN       <1     10       KRINGLE 3 (BY SIMILARITY).
FT   DOMAIN       34    112       KRINGLE 4 (BY SIMILARITY).
FT   DOMAIN      139   >169       KRINGLE 5 (BY SIMILARITY).
FT   DISULFID     34    112       BY SIMILARITY.
FT   DISULFID     55     95       BY SIMILARITY.
FT   DISULFID     83    107       BY SIMILARITY.
FT   NON_TER     169    169
SQ   SEQUENCE   169 AA;  18401 MW;  B55004D7 CRC32;

  Query Match           10.8%;  Score 204;  DB 1;  Length 169;
  Best Local Similarity  26.2%;  Pred. No. 1.10e-20;
  Matches   33;  Conservative   44;  Mismatches  40;  Indels   9;  Gaps   8;

Db      34 CYQGNGKSYRGTSSTTNTGKKCQSWVSMTPHSHSKTPANFPDSGLEMNYCRNPDNDQRGP 93
           |:  ||  ||   ::   | :| :|  : ::| : ::|  : :| : :|||||:| |||
Qy      25 CFWDNGHLYREDQTSPAPGLRCLNW--LDAQS-GLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db      94 WCFTT-DPSVR-WEYCNLKRCSETGG-GVAE-SAIVPQVPSAPGTSETDCMYGNGKEYRG 149
           ||: : :::|     |:  ||:|| : :::  :: : : : :||: | : ::: ::   :
Qy      81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQ-VFAPANALPA 139

Db     150 KTAVTA 155
           ::  :|
Qy     140 RSEAAA 145


RESULT    2
ID   PLMN_MOUSE     STANDARD;      PRT;   812 AA.
AC   P20918;
DT   01-FEB-1991 (REL. 17, CREATED)
DT   01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   PLASMINOGEN PRECURSOR (EC 3.4.21.7).
GN   PLG.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91184812.
RA   DEGEN S.J., BELL S.M., SCHAEFER L.A., ELLIOTT R.W.;
RT   "Characterization of the cDNA coding for mouse plasminogen and
RT   localization of the gene to mouse chromosome 17.";
RL   GENOMICS 8:49-61(1990).
CC   -!- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
CC       A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
CC       EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION,
CC       AND INFLAMMATION; IN OVULATION IT WEAKENS THE WALLS OF THE
CC       GRAAFIAN FOLLICLE. IT ACTIVATES THE UROKINASE-TYPE PLASMINOGEN
CC       ACTIVATOR, COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH
CC       AS C1 AND C5. IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
CC       LAMININ AND VON WILLEBRAND FACTOR.
CC   -!- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN
CC       ACTIVATORS, BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
CC       FIBRIN. CANNOT BE ACTIVATED WITH STREPTOKINASE.
CC   -!- PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN IMMEDIATELY AFTER
CC       DISSOCIATION FROM THE CLOT.
CC   -!- IN THE PRESENCE OF THE INHIBITOR, THE ACTIVATION INVOLVES ONLY
CC       CLEAVAGE AFTER ARG-581, RESULTING IN 2 CHAINS HELD TOGETHER BY 2
CC       DISULFIDE BONDS. WITHOUT THE INHIBITOR, THE ACTIVATION INVOLVES
CC       ALSO REMOVAL OF THE ACTIVATION PEPTIDE.
CC   -!- SIMILARITY: CONTAINS 5 KRINGLE REGIONS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04766; G200403; -.
DR   PIR; A38514; A38514.
DR   MGD; MGI:97620; PLG.
DR   PROSITE; PS00021; KRINGLE_1; 4.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PROSITE; PS50070; KRINGLE_2; 5.
DR   PFAM; PF00051; kringle; 5.
DR   PFAM; PF00089; trypsin; 1.
DR   HSSP; P00747; 1PMK.
KW   HYDROLASE; SERINE PROTEASE; PLASMA; GLYCOPROTEIN; FIBRINOLYSIS;
KW   TISSUE REMODELING; BLOOD COAGULATION; KRINGLE; ZYMOGEN; SIGNAL.
FT   SIGNAL        1     19
FT   CHAIN        20    812       PLASMINOGEN.
FT   CHAIN        20    581       HEAVY CHAIN A.
FT   PEPTIDE      20     96       ACTIVATION PEPTIDE.
FT   CHAIN        97    581       SHORT FORM OF CHAIN A.
FT   CHAIN       582    812       LIGHT CHAIN B.
FT   DOMAIN      103    181       KRINGLE 1.
FT   DOMAIN      184    262       KRINGLE 2.
FT   DOMAIN      275    352       KRINGLE 3.
FT   DOMAIN      377    454       KRINGLE 4.
FT   DOMAIN      481    560       KRINGLE 5.
FT   DOMAIN      582    812       CATALYTIC.
FT   ACT_SITE    624    624       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    667    667       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    762    762       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID     49     73       BY SIMILARITY.
FT   DISULFID     53     61       BY SIMILARITY.
FT   DISULFID    103    181       BY SIMILARITY.
FT   DISULFID    124    164       BY SIMILARITY.
FT   DISULFID    152    176       BY SIMILARITY.
FT   DISULFID    185    262       BY SIMILARITY.
FT   DISULFID    188    316       BY SIMILARITY.
FT   DISULFID    206    245       BY SIMILARITY.
FT   DISULFID    234    257       BY SIMILARITY.
FT   DISULFID    275    352       BY SIMILARITY.
FT   DISULFID    296    335       BY SIMILARITY.
FT   DISULFID    324    347       BY SIMILARITY.
FT   DISULFID    377    454       BY SIMILARITY.
FT   DISULFID    398    437       BY SIMILARITY.
FT   DISULFID    426    449       BY SIMILARITY.
FT   DISULFID    481    560       BY SIMILARITY.
FT   DISULFID    502    543       BY SIMILARITY.
FT   DISULFID    531    555       BY SIMILARITY.
FT   DISULFID    568    687       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    578    586       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    609    625       BY SIMILARITY.
FT   DISULFID    701    768       BY SIMILARITY.
FT   DISULFID    731    747       BY SIMILARITY.
FT   DISULFID    758    786       BY SIMILARITY.
SQ   SEQUENCE   812 AA;  90846 MW;  6C120F7A CRC32;

  Query Match           9.8%;  Score 184;  DB 1;  Length 812;
  Best Local Similarity  27.8%;  Pred. No. 1.20e-16;
  Matches   35;  Conservative   38;  Mismatches  43;  Indels  10;  Gaps   9;

Db     377 CYQSDGQSYRGTSSTTITGKKCQSWAAMFPHRHSKTPENFPDAGLEMNYCRNPDGD-KGP 435
           |:  :|: ||   ::   | :| :|    ::  : ::   : || : :|||||| | :||
Qy      25 CFWDNGHLYREDQTSPAPGLRCLNWLD--AQS-GLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db     436 WCYTT-DPSVR-WEYCNLKRCSETGG-SVVELPTVSQEPS-GPSDSETDCMYGNGKDYRG 491
           ||| : :::|     |:  ||:|| : ::  : |  ||:| ||:  | : ::: ::   :
Qy      81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQ-VFAPANALPA 139
```

```
Db    492 KTAVTA 497
          ::  :|
Qy    140 RSEAAA 145


RESULT    3
ID   APOA_HUMAN     STANDARD;      PRT;  4548 AA.
AC   P08519;
DT   01-AUG-1988 (REL. 08, CREATED)
DT   01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   APOLIPOPROTEIN(A) PRECURSOR (EC 3.4.21.-) (APO(A)) (LP(A)).
GN   LPA.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88039109.
RA   MCLEAN J.W., TOMLISON J.E., KUANG W.-J., EATON D.L., CHEN E.Y.,
RA   FLESS G.M., SCANU A.M., LAWN R.M.;
RT   "cDNA sequence of human apolipoprotein(a) is homologous to
RT   plasminogen.";
RL   NATURE 330:132-137(1987).
RN   [2]
RP   SERINE PROTEASE ACTIVITY.
RX   MEDLINE; 90076123.
RA   SALONEN E.-M., JAUHIAINEN M., ZARDI L., VAHERI A., EHNHOLM C.;
RT   "Lipoprotein(a) binds to fibronectin and has serine proteinase
RT   activity capable of cleaving it.";
RL   EMBO J. 8:4035-4040(1989).
RN   [3]
RP   REVIEW.
RX   MEDLINE; 90049223.
RA   UTERMANN G.;
RT   "The mysteries of lipoprotein(a).";
RL   SCIENCE 246:904-910(1989).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 4121-4208.
RX   MEDLINE; 96217891.
RA   MIKOL V., LOGRASSO P.V., BOETTCHER B.R.;
RT   "Crystal structures of apolipoprotein(a) kringle IV37 free and
RT   complexed with 6-aminohexanoic acid and with p-aminomethylbenzoic
RT   acid: existence of novel and expected binding modes.";
RL   J. MOL. BIOL. 256:751-761(1996).
RN   [5]
RP   VARIANT ARG-4193.
RX   MEDLINE; 95002201.
RA   SCANU A.M., PFAFFINGER D., LEE J.C., HINMAN J.;
RT   "A single point mutation (Trp72-->Arg) in human apo(a) kringle 4-37
RT   associated with a lysine binding defect in Lp(a).";
RL   BIOCHIM. BIOPHYS. ACTA 1227:41-45(1994).
CC   -!- FUNCTION: APO(A) IS THE MAIN CONSTITUENT OF LIPOPROTEIN(A). APO(A)
CC       CAN BIND TO FIBRONECTIN AND HAS SERINE PROTEINASE ACTIVITY CAPABLE
CC       OF CLEAVING IT.
CC   -!- SUBUNIT: APO(A) IS DISULFIDE LINKED TO APO-B100.
CC   -!- DISEASE: CONCENTRATION OF LP(A) IN PLASMA IS CORRELATED WITH
CC       ATHEROSCLEROSIS.
CC   -!- SIMILARITY: CONTAINS 38 KRINGLE REGIONS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X06290; G28620; -.
DR   PIR; S00657; S00657.
DR   MIM; 152200; -.
DR   PROSITE; PS00021; KRINGLE_1; 38.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PROSITE; PS50070; KRINGLE_2; 38.
DR   PFAM; PF00051; kringle; 38.
DR   PFAM; PF00089; trypsin; 1.
DR   HSSP; P00747; 1PMK.
KW   HYDROLASE; SERINE PROTEASE; LIPID TRANSPORT; PLASMA; GLYCOPROTEIN;
KW   KRINGLE; DUPLICATION; REPEAT; ATHEROSCLEROSIS; SIGNAL; POLYMORPHISM.
FT   SIGNAL       1     19
FT   CHAIN       20   4548       APOLIPOPROTEIN(A).
FT   DOMAIN      20    130       KRINGLE TYPE IV, 1.
FT   DOMAIN     131    244       KRINGLE TYPE IV, 2.
FT   DOMAIN     245    358       KRINGLE TYPE IV, 3.
FT   DOMAIN     359    472       KRINGLE TYPE IV, 4.
FT   DOMAIN     473    586       KRINGLE TYPE IV, 5.
FT   DOMAIN     587    700       KRINGLE TYPE IV, 6.
FT   DOMAIN     701    814       KRINGLE TYPE IV, 7.
FT   DOMAIN     815    928       KRINGLE TYPE IV, 8.
FT   DOMAIN     929   1042       KRINGLE TYPE IV, 9.
FT   DOMAIN    1043   1156       KRINGLE TYPE IV, 10.
FT   DOMAIN    1157   1270       KRINGLE TYPE IV, 11.
FT   DOMAIN    1271   1384       KRINGLE TYPE IV, 12.
FT   DOMAIN    1385   1498       KRINGLE TYPE IV, 13.
FT   DOMAIN    1499   1612       KRINGLE TYPE IV, 14.
FT   DOMAIN    1613   1726       KRINGLE TYPE IV, 15.
FT   DOMAIN    1727   1840       KRINGLE TYPE IV, 16.
FT   DOMAIN    1841   1954       KRINGLE TYPE IV, 17.
FT   DOMAIN    1955   2068       KRINGLE TYPE IV, 18.
FT   DOMAIN    2069   2182       KRINGLE TYPE IV, 19.
FT   DOMAIN    2183   2296       KRINGLE TYPE IV, 20.
FT   DOMAIN    2297   2410       KRINGLE TYPE IV, 21.
FT   DOMAIN    2411   2524       KRINGLE TYPE IV, 22.
FT   DOMAIN    2525   2638       KRINGLE TYPE IV, 23.
FT   DOMAIN    2639   2752       KRINGLE TYPE IV, 24.
FT   DOMAIN    2753   2866       KRINGLE TYPE IV, 25.
FT   DOMAIN    2867   2980       KRINGLE TYPE IV, 26.
FT   DOMAIN    2981   3094       KRINGLE TYPE IV, 27.
FT   DOMAIN    3095   3208       KRINGLE TYPE IV, 28.
FT   DOMAIN    3209   3322       KRINGLE TYPE IV, 29.
FT   DOMAIN    3323   3436       KRINGLE TYPE IV, 30.
FT   DOMAIN    3437   3550       KRINGLE TYPE IV, 31.
FT   DOMAIN    3551   3664       KRINGLE TYPE IV, 32.
FT   DOMAIN    3665   3770       KRINGLE TYPE IV, 33.
FT   DOMAIN    3771   3884       KRINGLE TYPE IV, 34.
FT   DOMAIN    3885   3998       KRINGLE TYPE IV, 35.
FT   DOMAIN    3999   4112       KRINGLE TYPE IV, 36.
FT   DOMAIN    4113   4226       KRINGLE TYPE IV, 37.
FT   DOMAIN    4227   4327       KRINGLE TYPE V.
FT   DOMAIN    4328   4548       CATALYTIC.
FT   ACT_SITE  4369   4369       CHARGE RELAY SYSTEM.
FT   ACT_SITE  4412   4412       CHARGE RELAY SYSTEM.
FT   ACT_SITE  4498   4498       CHARGE RELAY SYSTEM.
FT   VARIANT   4193   4193       W -> R (LOSS OF LYSINE-SEPHAROSE
FT                               BINDING).
SQ   SEQUENCE   4548 AA;  501313 MW;  51F6509C CRC32;

  Query Match           9.7%;  Score 183;  DB 1;  Length 4548;
  Best Local Similarity  27.5%;  Pred. No. 1.89e-16;
  Matches    36;  Conservative   34;  Mismatches  52;  Indels   9;  Gaps   8;

Db   3539 ILAPSLEAFF-EQALTEETPGVQDCYYHYGQSYRGTYSTTVTGRTCQAWSSMTPHQHSRT 3597
          :| : ::||:    |  |: |   |:   |: ||   ::   |  |  |  : : | : :
Qy      1 MLLAWVQAFLVSNMLLAEAYGSGGCFWDNGHLYREDQTSPAPGLRCLNW--LDA-QSGLA 57

Db   3598 PENYPNAGLTRNYCRNPDAEIR-PWCYTM-DPSVR-WEYCNLTQCLVTESSVLATLTV-V 3653
           :  : ||  ::|||||| : | ||||   :::|     |:  :|  | |  |:::|  :
Qy     58 SAPVSGAG-NHSYCRNPDEDPRGPWCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEI 116

Db   3654 PDPSTEASSEE 3664
           ::|  ::::|
Qy    117 QEASEGPGADE 127
```

```
RESULT    4
ID   HGFA_HUMAN     STANDARD;      PRT;   655 AA.
AC   Q04756; Q14726;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   HEPATOCYTE GROWTH FACTOR ACTIVATOR PRECURSOR (EC 3.4.21.-) (HGF
DE   ACTIVATOR).
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=LIVER, AND SERUM;
RX   MEDLINE; 93252878.
RA   MIYAZAWA K., SHIMOMURA T., KITAMURA A., KONDO J., MORIMOTO Y.,
RA   KITAMURA N.;
RT   "Molecular cloning and sequence analysis of the cDNA for a human
RT   serine protease reponsible for activation of hepatocyte growth
RT   factor. Structural similarity of the protease precursor to blood
RT   coagulation factor XII.";
RL   J. BIOL. CHEM. 268:10024-10028(1993).
RN   [2]
RP   SEQUENCE OF 40-655 FROM N.A.
RA   ZHAO S., ODELL C.;
RL   SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: ACTIVATES HEPATOCYTE GROWTH FACTOR (HGF) BY
CC       CONVERTING IT FROM A SINGLE CHAIN TO A HETERODIMERIC FORM.
CC   -!- SUBUNIT: DIMER OF A SHORT CHAIN AND A LONG CHAIN LINKED BY A
CC       DISULFIDE BOND.
CC   -!- SUBCELLULAR LOCATION: SECRETED AS AN INACTIVE SINGLE-CHAIN
CC       PRECURSOR AND IS THEN ACTIVATED TO A HETERODIMERIC FORM.
CC   -!- TISSUE SPECIFICITY: LIVER.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE-I DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE-II DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 KRINGLE REGION.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 IS THE INITIATOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D14012; G219681; -.
DR   EMBL; Z69923; E225802; -.
DR   PIR; A46688; A46688.
DR   PROSITE; PS00021; KRINGLE_1; 1.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS00023; FIBRONECTIN_2; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01253; FIBRONECTIN_1; 1.
DR   PROSITE; PS50070; KRINGLE_2; 1.
DR   PFAM; PF00008; EGF; 2.
DR   PFAM; PF00039; fn1; 1.
DR   PFAM; PF00040; fn2; 1.
DR   PFAM; PF00051; kringle; 1.
DR   PFAM; PF00089; trypsin; 1.
DR   HSSP; P00763; 1DPO.
KW   HYDROLASE; GLYCOPROTEIN; PLASMA; SERINE PROTEASE; KRINGLE; SIGNAL;
KW   EGF-LIKE DOMAIN; REPEAT; ZYMOGEN.
FT   SIGNAL        1     30
FT   PROPEP       31    372       CLEAVED IN ACTIVE FORM.
FT   CHAIN       373    407       HEPATOCYTE GROWTH FACTOR ACTIVATOR SHORT
FT                                CHAIN.
FT   CHAIN       408    655       HEPATOCYTE GROWTH FACTOR ACTIVATOR LONG
FT                                CHAIN.
FT   DOMAIN      108    148       FIBRONECTIN TYPE-II.
FT   DOMAIN      160    198       EGF-LIKE 1.
FT   DOMAIN      200    240       FIBRONECTIN TYPE-I.
FT   DOMAIN      241    279       EGF-LIKE 2.
FT   DOMAIN      286    367       KRINGLE.
FT   DOMAIN      408    655       CATALYTIC.
FT   ACT_SITE    447    447       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    497    497       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    598    598       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID    108    133       BY SIMILARITY.
FT   DISULFID    122    148       BY SIMILARITY.
FT   DISULFID    164    175       BY SIMILARITY.
FT   DISULFID    169    186       BY SIMILARITY.
FT   DISULFID    188    197       BY SIMILARITY.
FT   DISULFID    202    230       BY SIMILARITY.
FT   DISULFID    228    237       BY SIMILARITY.
FT   DISULFID    245    256       BY SIMILARITY.
FT   DISULFID    250    267       BY SIMILARITY.
FT   DISULFID    269    278       BY SIMILARITY.
FT   DISULFID    286    367       BY SIMILARITY.
FT   DISULFID    307    349       BY SIMILARITY.
FT   DISULFID    338    362       BY SIMILARITY.
FT   DISULFID    394    521       INTRACHAIN (BY SIMILARITY).
FT   DISULFID    432    448       BY SIMILARITY.
FT   DISULFID    440    510       BY SIMILARITY.
FT   DISULFID    535    604       BY SIMILARITY.
FT   DISULFID    567    583       BY SIMILARITY.
FT   DISULFID    594    622       BY SIMILARITY.
FT   CARBOHYD     48     48       POTENTIAL.
FT   CARBOHYD    290    290       POTENTIAL.
FT   CARBOHYD    468    468       POTENTIAL.
FT   CARBOHYD    492    492       POTENTIAL.
FT   CARBOHYD    546    546       POTENTIAL.
FT   CONFLICT    644    644       R -> Q (IN REF. 2).
SQ   SEQUENCE   655 AA;  70681 MW;  BFE0842D CRC32;

  Query Match             9.6%;  Score 181;  DB 1;  Length 655;
  Best Local Similarity  40.5%;  Pred. No. 4.70e-16;
  Matches    30;  Conservative   19;  Mismatches  17;  Indels   8;  Gaps   4;

Db    286 CFLGNGTGYRGVASTSASGLSCLAWNSDLLYQELHVDSVGAAALLGLGPHAYCRNPDNDE 345
            ||  ||  ||   :::|:|| || |    |  : :  ::::|:: | | |:||||||:|
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW----L--DAQ-SGLASAPVSGAGNHSYCRNPDEDP 77

Db    346 R-PWCYVVKDSALS 358
            | |||||  :::::
Qy     78 RGPWCYVSGEAGVP 91


RESULT    5
ID   HGFL_MOUSE     STANDARD;      PRT;   716 AA.
AC   P26928;
DT   01-AUG-1992 (REL. 23, CREATED)
DT   01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   HEPATOCYTE GROWTH FACTOR-LIKE PROTEIN PRECURSOR (MACROPHAGE
DE   STIMULATORY PROTEIN) (MSP).
GN   HGFL.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=LIVER;
RX   MEDLINE; 92002017.
RA   FRIEZNER DEGEN S.J., STUART L.A., HAN S., JAMISON C.S.;
RT   "Characterization of the mouse cDNA and gene coding for a hepatocyte
RT   growth factor-like protein: expression during development.";
RL   BIOCHEMISTRY 30:9781-9791(1991).
CC   -!- FUNCTION: PROBABLY HAS NO PROTEOLYTIC ACTIVITY, SINCE CRUCIAL AA
CC       CHARACTERISTIC OF SERINE PROTEASES CATALYTIC SITES ARE NOT
```

```
CC       CONSERVED.
CC    -!- TISSUE SPECIFICITY: LIVER. LOWER LEVELS IN LUNG, PLACENTA AND
CC       ADRENAL.
CC    -!- DEVELOPMENTAL STAGE: IS EXPRESSED AT LOW LEVELS DURING GESTATION.
CC       JUST BEFORE BIRTH THE LEVEL INCREASES DRAMATICALLY AND REMAINS
CC       STABLE AFTERWARDS.
CC    -!- PTM: MAY BE CLEAVED AFTER AA 488, TO YIELD A TWO-CHAIN MOLECULE
CC       HELD TOGETHER BY DISULFIDE BONDS, OR CLEAVED INTO TWO SEPARATE
CC       POLYPEPTIDES.
CC    -!- SIMILARITY: CONTAINS 4 KRINGLE REGIONS.
CC    -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
CC    ---------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    ---------------------------------------------------------------------
DR    EMBL; M74180; G193832; -.
DR    EMBL; M74181; G193834; -.
DR    MGD; MGI:96080; HGFL.
DR    PROSITE; PS00021; KRINGLE_1; 4.
DR    PROSITE; PS50070; KRINGLE_2; 4.
DR    PFAM; PF00051; kringle; 4.
DR    PFAM; PF00089; trypsin; 1.
DR    HSSP; P00747; 1PMK.
KW    KRINGLE; GLYCOPROTEIN; SERINE PROTEASE HOMOLOG; SIGNAL.
FT    SIGNAL        1     31       POTENTIAL.
FT    CHAIN        32    716       HEPATOCYTE GROWTH FACTOR-LIKE PROTEIN.
FT    DOMAIN       32    109       PAP.
FT    DOMAIN      110    186       KRINGLE 1.
FT    DOMAIN      191    268       KRINGLE 2.
FT    DOMAIN      292    370       KRINGLE 3.
FT    DOMAIN      379    457       KRINGLE 4.
FT    DOMAIN      489    716       SERINE PROTEASE-LIKE.
FT    DISULFID     56     78       BY SIMILARITY.
FT    DISULFID     60     66       BY SIMILARITY.
FT    DISULFID    110    186       BY SIMILARITY.
FT    DISULFID    131    169       BY SIMILARITY.
FT    DISULFID    157    181       BY SIMILARITY.
FT    DISULFID    191    268       BY SIMILARITY.
FT    DISULFID    194    333       INTERCHAIN (BY SIMILARITY).
FT    DISULFID    212    251       BY SIMILARITY.
FT    DISULFID    240    263       BY SIMILARITY.
FT    DISULFID    292    370       BY SIMILARITY.
FT    DISULFID    313    352       BY SIMILARITY.
FT    DISULFID    341    364       BY SIMILARITY.
FT    DISULFID    379    457       BY SIMILARITY.
FT    DISULFID    400    440       BY SIMILARITY.
FT    DISULFID    428    452       BY SIMILARITY.
FT    DISULFID    477    593       INTERCHAIN (BY SIMILARITY).
FT    DISULFID    512    528       BY SIMILARITY.
FT    DISULFID    607    672       BY SIMILARITY.
FT    DISULFID    637    651       BY SIMILARITY.
FT    DISULFID    662    690       BY SIMILARITY.
FT    CARBOHYD     72     72       POTENTIAL.
FT    CARBOHYD    173    173       POTENTIAL.
FT    CARBOHYD    305    305       POTENTIAL.
FT    CARBOHYD    620    620       POTENTIAL.
FT    CONFLICT     19     19       P -> Q (IN GENOMIC SEQUENCE).
SQ    SEQUENCE   716 AA;  80588 MW;  EDDAD87F CRC32;

  Query Match             9.4%;  Score 177;  DB 1;  Length 716;
  Best Local Similarity  41.3%;  Pred. No. 2.88e-15;
  Matches    26;  Conservative    9;  Mismatches  26;  Indels   2;  Gaps   2;

Db    110 CIMDNGVSYRGTVARTAGGLPCQAWSRRFPNDHKYTPTPKNGLEENFCRNPDGDPRGPWC 169
        |: |||  ||   :  | || |  |     :    :| :  | : ::||||| |||||||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLDA-QSGLASAPVSGAG-NHSYCRNPDEDPRGPWC 82

Db    170 YTT 172
          | :
Qy     83 YVS 85


RESULT    6
ID    PLMN_ERIEU     STANDARD;      PRT;   810 AA.
AC    Q29485;
DT    01-NOV-1997 (REL. 35, CREATED)
DT    01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT    01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE    PLASMINOGEN PRECURSOR (EC 3.4.21.7).
GN    PLG.
OS    ERINACEUS EUROPAEUS (WESTERN EUROPEAN HEDGEHOG).
OC    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC    INSECTIVORA; ERINACEIDAE; ERINACEINAE; ERINACEUS.
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=LIVER;
RX    MEDLINE; 96025778.
RA    LAWN R.M., BOONMARK N.W., SCHWARTZ K., LINDAHL G.E., WADE D.P.,
RA    BYRNE C.D., FONG K.J., MEER K., PATTHY L.;
RT    "The recurring evolution of lipoprotein(a). Insights from cloning of
RT    hedgehog apolipoprotein(a).";
RL    J. BIOL. CHEM. 270:24004-24009(1995).
RN    [2]
RP    REVISIONS.
RA    LAWN R.M.;
RL    SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC    -!- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
CC       A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
CC       EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION,
CC       AND INFLAMMATION; IN OVULATION IT WEAKENS THE WALLS OF THE
CC       GRAAFIAN FOLLICLE. IT ACTIVATES THE UROKINASE-TYPE PLASMINOGEN
CC       ACTIVATOR, COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH
CC       AS C1 AND C5. IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
CC       LAMININ AND VON WILLEBRAND FACTOR.
CC    -!- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN
CC       ACTIVATORS, BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
CC       FIBRIN. CANNOT BE ACTIVATED WITH STREPTOKINASE.
CC    -!- PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN IMMEDIATELY AFTER
CC       DISSOCIATION FROM THE CLOT.
CC    -!- SIMILARITY: CONTAINS 5 KRINGLE REGIONS.
CC    -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
CC    ---------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    ---------------------------------------------------------------------
DR    EMBL; U33171; G1806583; -.
DR    PROSITE; PS00021; KRINGLE_1; 5.
DR    PROSITE; PS00134; TRYPSIN_HIS; 1.
DR    PROSITE; PS00135; TRYPSIN_SER; 1.
DR    PROSITE; PS50070; KRINGLE_2; 5.
DR    PFAM; PF00051; kringle; 5.
DR    PFAM; PF00089; trypsin; 1.
DR    HSSP; P00747; 1PMK.
KW    HYDROLASE; SERINE PROTEASE; PLASMA; GLYCOPROTEIN; FIBRINOLYSIS;
KW    TISSUE REMODELING; BLOOD COAGULATION; KRINGLE; ZYMOGEN; SIGNAL.
FT    SIGNAL        1     19       BY SIMILARITY.
FT    CHAIN        20    810       PLASMINOGEN.
FT    CHAIN        20    582       HEAVY CHAIN A (BY SIMILARITY).
FT    CHAIN       583    810       LIGHT CHAIN B (BY SIMILARITY).
FT    DOMAIN      583    810       CATALYTIC.
FT    DOMAIN      103    181       KRINGLE 1.
FT    DOMAIN      185    262       KRINGLE 2.
FT    DOMAIN      275    352       KRINGLE 3.
FT    DOMAIN      379    456       KRINGLE 4.
```

```
FT   DOMAIN      482    561       KRINGLE 5.
FT   ACT_SITE    622    622       CHARGE RELAY SYSTEM.
FT   ACT_SITE    665    665       CHARGE RELAY SYSTEM.
FT   ACT_SITE    760    760       CHARGE RELAY SYSTEM.
FT   CARBOHYD    339    339       POTENTIAL.
SQ   SEQUENCE   810 AA;  90902 MW;  ABFDC531 CRC32;

  Query Match             9.3%;  Score 176;  DB 1;  Length 810;
  Best Local Similarity  32.9%;  Pred. No. 4.53e-15;
  Matches    26;  Conservative   20;  Mismatches  27;  Indels   6;  Gaps   6;

Db    107 NGKYYRGTVSKTKTGLTCQKWSAETPHKPRFSPDENPSEGLDQNYCRNPDNDPKGPWCYT 166
          ||  ||   :    || | :|  :: :   :::   :: | :::||||||:||:|||||
Qy     29 NGHLYREDQTSPAPGLRCLNW-LDA-QSG-LASAPVSGAG-NHSYCRNPDEDPRGPWCYV 84

Db    167 M-DPEVRYEY-CEIIQCED 183
           :: |    || ::| :
Qy     85 SGEAGVPEKRPCEDLRCPE 103


RESULT    7
ID   PLMN_MACMU     STANDARD;      PRT;   810 AA.
AC   P12545;
DT   01-OCT-1989 (REL. 12, CREATED)
DT   01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   PLASMINOGEN PRECURSOR (EC 3.4.21.7).
GN   PLG.
OS   MACACA MULATTA (RHESUS MACAQUE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; MACACA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89174660.
RA   TOMLINSON J.E., MCLEAN J.W., LAWN R.M.;
RT   "Rhesus monkey apolipoprotein(a). Sequence, evolution, and sites of
RT   synthesis.";
RL   J. BIOL. CHEM. 264:5957-5965(1989).
CC   -!- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
CC       A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
CC       EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION,
CC       AND INFLAMMATION; IN OVULATION IT WEAKENS THE WALLS OF THE
CC       GRAAFIAN FOLLICLE. IT ACTIVATES THE UROKINASE-TYPE PLASMINOGEN
CC       ACTIVATOR, COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH
CC       AS C1 AND C5. IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
CC       LAMININ AND VON WILLEBRAND FACTOR.
CC   -!- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN
CC       ACTIVATORS, BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
CC       FIBRIN. ACTIVATED WITH CATALYTIC AMOUNTS OF STREPTOKINASE.
CC   -!- PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN IMMEDIATELY AFTER
CC       DISSOCIATION FROM THE CLOT.
CC   -!- IN THE PRESENCE OF THE INHIBITOR, THE ACTIVATION INVOLVES ONLY
CC       CLEAVAGE AFTER ARG-580, RESULTING IN 2 CHAINS HELD TOGETHER BY 2
CC       DISULFIDE BONDS. WITHOUT THE INHIBITOR, THE ACTIVATION INVOLVES
CC       ALSO REMOVAL OF THE ACTIVATION PEPTIDE.
CC   -!- SIMILARITY: CONTAINS 5 KRINGLE REGIONS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04697; G342273; -.
DR   PIR; B30848; B30848.
DR   PIR; B32869; B32869.
DR   PROSITE; PS00021; KRINGLE_1; 5.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PROSITE; PS50070; KRINGLE_2; 5.
DR   PFAM; PF00051; kringle; 5.
DR   PFAM; PF00089; trypsin; 1.
DR   HSSP; P00747; 1PMK.
KW   HYDROLASE; SERINE PROTEASE; PLASMA; GLYCOPROTEIN; FIBRINOLYSIS;
KW   TISSUE REMODELING; BLOOD COAGULATION; KRINGLE; ZYMOGEN; SIGNAL.
FT   SIGNAL        1     19
FT   CHAIN        20    810       PLASMINOGEN.
FT   CHAIN        20    580       HEAVY CHAIN A.
FT   PEPTIDE      20     96       ACTIVATION PEPTIDE.
FT   CHAIN        97    580       SHORT FORM OF CHAIN A.
FT   CHAIN       581    810       LIGHT CHAIN B.
FT   DOMAIN      103    181       KRINGLE 1.
FT   DOMAIN      184    262       KRINGLE 2.
FT   DOMAIN      275    352       KRINGLE 3.
FT   DOMAIN      377    454       KRINGLE 4.
FT   DOMAIN      481    560       KRINGLE 5.
FT   DOMAIN      581    810       CATALYTIC.
FT   ACT_SITE    622    622       CHARGE RELAY SYSTEM.
FT   ACT_SITE    665    665       CHARGE RELAY SYSTEM.
FT   ACT_SITE    760    760       CHARGE RELAY SYSTEM.
FT   BINDING     136    136       OMEGA-AMINOCARBOXYLIC ACIDS.
FT   BINDING     158    158       OMEGA-AMINOCARBOXYLIC ACIDS.
FT   BINDING     172    172       OMEGA-AMINOCARBOXYLIC ACIDS.
FT   BINDING     432    432       OMEGA-AMINOCARBOXYLIC ACIDS.
FT   BINDING     445    445       OMEGA-AMINOCARBOXYLIC ACIDS.
FT   BINDING     134    134       FIBRIN.
FT   BINDING     136    136       FIBRIN.
FT   DISULFID     49     73       BY SIMILARITY.
FT   DISULFID     53     61       BY SIMILARITY.
FT   DISULFID    103    181       BY SIMILARITY.
FT   DISULFID    124    164       BY SIMILARITY.
FT   DISULFID    152    176       BY SIMILARITY.
FT   DISULFID    185    262       BY SIMILARITY.
FT   DISULFID    188    316       BY SIMILARITY.
FT   DISULFID    206    245       BY SIMILARITY.
FT   DISULFID    234    257       BY SIMILARITY.
FT   DISULFID    275    352       BY SIMILARITY.
FT   DISULFID    296    335       BY SIMILARITY.
FT   DISULFID    324    347       BY SIMILARITY.
FT   DISULFID    377    454       BY SIMILARITY.
FT   DISULFID    398    437       BY SIMILARITY.
FT   DISULFID    426    449       BY SIMILARITY.
FT   DISULFID    481    560       BY SIMILARITY.
FT   DISULFID    502    543       BY SIMILARITY.
FT   DISULFID    531    555       BY SIMILARITY.
FT   DISULFID    567    685       BY SIMILARITY.
FT   DISULFID    577    585       BY SIMILARITY.
FT   DISULFID    607    623       BY SIMILARITY.
FT   DISULFID    699    766       BY SIMILARITY.
FT   DISULFID    729    745       BY SIMILARITY.
FT   DISULFID    756    784       BY SIMILARITY.
FT   CARBOHYD    365    365       BY SIMILARITY.
SQ   SEQUENCE   810 AA;  90255 MW;  B5AA0F07 CRC32;

  Query Match             9.2%;  Score 173;  DB 1;  Length 810;
  Best Local Similarity  24.8%;  Pred. No. 1.75e-14;
  Matches    31;  Conservative   41;  Mismatches  43;  Indels  10;  Gaps   9;

Db    377 CYHGDGQSYRGTSSTTTGKKCQSWSSMTPHWHEKTPENFPNAGLTMNYCRNPDAD-KGP 435
          |:  :|: ||   :: : | :| :|  : ::    ::   : ||   :|||||| | :||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW--LDAQ-SGLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db    436 WCFTT-DPSVR-WEYCNLKKCSGTEGSVAAPPPVAQLPDA-ETPSEEDC-MFGNGKGYRG 491
          ||: : :::|     |:  :|: |  | | |:  ::: :| | |: ::  :|: :::  :
Qy     81 WCYVSGEAGVPEKRPCEDLRCPET-TSQALPAFTTEIQEASEGPGADEVQVFAPANALPA 139

Db    492 KKATT 496
          :  ::
Qy    140 RSEAA 144


RESULT    8
```

```
ID   PLMN_PIG       STANDARD;      PRT;   790 AA.
AC   P06867;
DT   01-JAN-1988 (REL. 06, CREATED)
DT   01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   PLASMINOGEN (EC 3.4.21.7).
OS   SUS SCROFA (PIG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN   [1]
RP   SEQUENCE OF 1-560.
RA   SCHALLER J., MARTI T., ROESSELET S.J., KAEMPFER U., RICKLI E.E.;
RT   "Amino acid sequence of the heavy chain of porcine plasmin. Comparison
RT   of the carbohydrate attachment sites with the human and bovine
RT   species.";
RL   FIBRINOLYSIS 1:91-102(1987).
RN   [2]
RP   SEQUENCE OF 450-790.
RX   MEDLINE; 85203907.
RA   MARTI T., SCHALLER J., RICKLI E.E.;
RT   "Determination of the complete amino-acid sequence of porcine
RT   miniplasminogen.";
RL   EUR. J. BIOCHEM. 149:279-285(1985).
CC   -!- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
CC       A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
CC       EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION,
CC       AND INFLAMMATION; IN OVULATION IT WEAKENS THE WALLS OF THE
CC       GRAAFIAN FOLLICLE. IT ACTIVATES THE UROKINASE-TYPE PLASMINOGEN
CC       ACTIVATOR, COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH
CC       AS C1 AND C5. IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
CC       LAMININ AND VON WILLEBRAND FACTOR.
CC   -!- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN
CC       ACTIVATORS, BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
CC       FIBRIN. CANNOT BE ACTIVATED WITH STREPTOKINASE.
CC   -!- PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN IMMEDIATELY AFTER
CC       DISSOCIATION FROM THE CLOT.
CC   -!- SIMILARITY: CONTAINS 5 KRINGLE REGIONS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
DR   PIR; A25834; A25834.
DR   PIR; S03733; S03733.
DR   PROSITE; PS00134; TRYPSIN_HIS; FALSE_NEG.
DR   PROSITE; PS00021; KRINGLE_1; 5.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PROSITE; PS50070; KRINGLE_2; 5.
DR   PFAM; PF00051; kringle; 5.
DR   PFAM; PF00089; trypsin; 1.
DR   HSSP; P00747; 5HPG.
KW   HYDROLASE; SERINE PROTEASE; PLASMA; GLYCOPROTEIN; FIBRINOLYSIS;
KW   TISSUE REMODELING; BLOOD COAGULATION; KRINGLE; ZYMOGEN.
FT   CHAIN         1    560       HEAVY CHAIN A.
FT   CHAIN       561    790       LIGHT CHAIN B.
FT   DOMAIN      561    790       CATALYTIC.
FT   DOMAIN       84    162       KRINGLE 1.
FT   DOMAIN      166    243       KRINGLE 2.
FT   DOMAIN      256    333       KRINGLE 3.
FT   DOMAIN      358    435       KRINGLE 4.
FT   DOMAIN      461    540       KRINGLE 5.
FT   ACT_SITE    602    602       CHARGE RELAY SYSTEM.
FT   ACT_SITE    645    645       CHARGE RELAY SYSTEM.
FT   ACT_SITE    740    740       CHARGE RELAY SYSTEM.
FT   CARBOHYD    249    249
SQ   SEQUENCE   790 AA;  88592 MW;  EE597814 CRC32;

  Query Match             9.1%;  Score 171;  DB 1;  Length 790;
  Best Local Similarity  26.4%;  Pred. No. 4.27e-14;
  Matches    34;  Conservative   36;  Mismatches  49;  Indels  10;  Gaps   9;

Db    358 CYRGNGESYRGTSSTTITGRKCQSWVSMTPHRHEKTPGNFPNAGLTMNYCRNPDAD-KSP 416
          |:: ||  ||   ::   | :| :|  : ::    :::  : ||   :|||||| | ::|
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW--LDAQS-GLASAPVSGAG-NHSYCRNPDEDPRGP 80

Db    417 WCYTT-DPRVR-WEYCNLKKCSET-EQQVTNFPAIAQVPSVEDLSEDC-MFGNGKRYRGK 472
          ||| : :: |     |:  :|:||  | :  | :  | :|    :::  :|: ::   ::
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQVFAPANALPAR 140

Db    473 R-ATTVAGV 480
           |::|  |
Qy    141 SEAAAVQPV 149


RESULT    9
ID   PLMN_BOVIN     STANDARD;      PRT;   812 AA.
AC   P06868; Q28162;
DT   01-JAN-1988 (REL. 06, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   PLASMINOGEN PRECURSOR (EC 3.4.21.7).
GN   PLG.
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   BERGLUND L., ANDERSEN M.D., PETERSEN T.E.;
RT   "Cloning and characterization of the bovine plasminogen cDNA.";
RL   INT. DAIRY J. 5:593-603(1995).
RN   [2]
RP   SEQUENCE OF 27-812, AND CARBOHYDRATE-BINDING SITES.
RX   MEDLINE; 85203906.
RA   SCHALLER J., MOSER P.W., DANNEGGER-MULLER G.A.K., ROSSELET S.J.,
RA   KAMPFER U., RICKLI E.E.;
RT   "Complete amino acid sequence of bovine plasminogen. Comparison with
RT   human plasminogen.";
RL   EUR. J. BIOCHEM. 149:267-278(1985).
RN   [3]
RP   SEQUENCE OF 706-812 FROM N.A.
RX   MEDLINE; 85023311.
RA   MALINOWSKI D.P., SADLER J.E., DAVIE E.W.;
RT   "Characterization of a complementary deoxyribonucleic acid coding for
RT   human and bovine plasminogen.";
RL   BIOCHEMISTRY 23:4243-4250(1984).
CC   -!- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
CC       A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
CC       EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION,
CC       AND INFLAMMATION; IN OVULATION IT WEAKENS THE WALLS OF THE
CC       GRAAFIAN FOLLICLE. IT ACTIVATES THE UROKINASE-TYPE PLASMINOGEN
CC       ACTIVATOR, COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH
CC       AS C1 AND C5. IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
CC       LAMININ AND VON WILLEBRAND FACTOR.
CC   -!- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN
CC       ACTIVATORS, BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
CC       FIBRIN. CANNOT BE ACTIVATED WITH STREPTOKINASE.
CC   -!- PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN IMMEDIATELY AFTER
CC       DISSOCIATION FROM THE CLOT.
CC   -!- SIMILARITY: CONTAINS 5 KRINGLE REGIONS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X79402; G494963; -.
DR   EMBL; K02935; G163552; -.
DR   PIR; A25835; PLBO.
DR   PROSITE; PS00021; KRINGLE_1; 5.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PROSITE; PS50070; KRINGLE_2; 5.
DR   PFAM; PF00051; kringle; 5.
```

```
DR   PFAM; PF00089; trypsin; 1.
DR   HSSP; P00747; 2PK4.
KW   HYDROLASE; SERINE PROTEASE; PLASMA; GLYCOPROTEIN; FIBRINOLYSIS;
KW   TISSUE REMODELING; BLOOD COAGULATION; KRINGLE; ZYMOGEN; SIGNAL.
FT   SIGNAL        1     26
FT   CHAIN        27    812       PLASMINOGEN.
FT   CHAIN        27    583       HEAVY CHAIN A.
FT   CHAIN       584    812       LIGHT CHAIN B.
FT   DOMAIN      584    812       CATALYTIC.
FT   DOMAIN      110    188       KRINGLE 1.
FT   DOMAIN      192    269       KRINGLE 2.
FT   DOMAIN      282    359       KRINGLE 3.
FT   DOMAIN      384    461       KRINGLE 4.
FT   DOMAIN      485    564       KRINGLE 5.
FT   CARBOHYD    315    315
FT   CARBOHYD    365    365
FT   ACT_SITE    624    624       CHARGE RELAY SYSTEM.
FT   ACT_SITE    667    667       CHARGE RELAY SYSTEM.
FT   ACT_SITE    762    762       CHARGE RELAY SYSTEM.
FT   CONFLICT    335    335       N -> D (IN REF. 2).
FT   CONFLICT    516    516       Q -> H (IN REF. 2).
FT   CONFLICT    555    555       P -> L (IN REF. 2).
FT   CONFLICT    744    744       T -> R (IN REF. 3).
SQ   SEQUENCE   812 AA;  91216 MW;  3E3C0328 CRC32;

  Query Match             9.1%;  Score 172;  DB 1;  Length 812;
  Best Local Similarity  31.1%;  Pred. No. 2.73e-14;
  Matches    33;  Conservative   25;  Mismatches  39;  Indels   9;  Gaps   8;

Db    384 CYHGNGQSYRGTSSTTITGRKCQSWSSMTPHRHLKTPENYPNAGLTMNYCRNPDAD-KSP 442
          |:  ||: ||   ::   | :| :|  : ::  | ::   : ||   :|||||| | ::|
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW--LDAQSGL-ASAPVSGAG-NHSYCRNPDEDPRGP 80

Db    443 WCYTT-DPRVR-WEFCNLKKCSETPEQ-VPA-APQAPGVENPPEAD 484
          ||| : :: |     |:  :|:||  | :|| : :     : | ||
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGAD 126


RESULT   10
ID   HGF_MOUSE      STANDARD;      PRT;   728 AA.
AC   Q08048;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   HEPATOCYTE GROWTH FACTOR PRECURSOR (SCATTER FACTOR) (SF)
DE   (HEPATOPOEITIN A).
GN   HGF.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 496-504.
RC   TISSUE=MAMMARY FIBROBLAST;
RX   MEDLINE; 94183257.
RA   SASAKI M., NISHIO M., SASAKI T., ENAMI J.;
RT   "Identification of mouse mammary fibroblast-derived mammary growth
RT   factor as hepatocyte growth factor.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 199:772-779(1994).
CC   -!- FUNCTION: HGF IS A POTENT MITOGEN FOR MATURE PARENCHYMAL
CC       HEPATOCYTE CELLS, SEEMS TO BE AN HEPATOTROPHIC FACTOR, AND ACTS
CC       AS GROWTH FACTOR FOR A BROAD SPECTRUM OF TISSUES AND CELL TYPES.
CC       IT HAS NO DETECTABLE PROTEASE ACTIVITY.
CC   -!- SUBUNIT: DIMER OF AN ALPHA CHAIN AND A BETA CHAIN LINKED BY A
CC       DISULFIDE BOND.
CC   -!- ALTERNATIVE PRODUCTS: A SHORT FORM OF HGF IS PRODUCED BY
CC       ALTERNATIVE RNA SPLICING. THE SEQUENCE SHOWN HERE IS THAT OF THE
CC       LONG FORM.
CC   -!- SIMILARITY: CONTAINS 4 KRINGLE REGIONS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10212; G220436; -.
DR   EMBL; D10213; G220438; -.
DR   MGD; MGI:96079; HGF.
DR   PROSITE; PS00021; KRINGLE_1; 4.
DR   PROSITE; PS50070; KRINGLE_2; 4.
DR   PFAM; PF00051; kringle; 4.
DR   PFAM; PF00089; trypsin; 1.
DR   HSSP; P14210; 2HGF.
KW   GROWTH FACTOR; KRINGLE; GLYCOPROTEIN; SERINE PROTEASE HOMOLOG;
KW   SIGNAL; ALTERNATIVE SPLICING.
FT   SIGNAL        1     32       BY SIMILARITY.
FT   CHAIN        33    495       HEPATOCYTE GROWTH FACTOR ALPHA CHAIN.
FT   CHAIN       496    728       HEPATOCYTE GROWTH FACTOR BETA CHAIN.
FT   MOD_RES      33     33       PYRROLIDONE CARBOXYLIC ACID
FT                                (BY SIMILARITY).
FT   DOMAIN       33    128       PAP.
FT   DOMAIN      129    207       KRINGLE 1.
FT   DOMAIN      212    289       KRINGLE 2.
FT   DOMAIN      306    384       KRINGLE 3.
FT   DOMAIN      392    470       KRINGLE 4.
FT   DOMAIN      496    728       SERINE PROTEASE-LIKE.
FT   DISULFID     71     97       BY SIMILARITY.
FT   DISULFID     75     85       BY SIMILARITY.
FT   DISULFID    488    607       INTERCHAIN (BY SIMILARITY).
FT   CARBOHYD    295    295       POTENTIAL.
FT   CARBOHYD    403    403       POTENTIAL.
FT   CARBOHYD    569    569       POTENTIAL.
FT   CARBOHYD    656    656       POTENTIAL.
FT   VARSPLIC    163    167       MISSING (IN SHORT FORM).
SQ   SEQUENCE   728 AA;  82944 MW;  967CDF1E CRC32;

  Query Match             8.9%;  Score 168;  DB 1;  Length 728;
  Best Local Similarity  31.9%;  Pred. No. 1.62e-13;
  Matches    30;  Conservative   23;  Mismatches  34;  Indels   7;  Gaps   5;

Db    387 SSGQDCYRGNGKNYMGNLSKTRSGLTCSMWDKNMEDLHRHIFWEPDASKLNKNYCRNPDD 446
          :||  |:: ||  |  : :   :|| |  |   ::  :  :   | ::  | :||||||:
Qy     21 GSG-GCFWDNGHLYREDQTSPAPGLRCLNW---LDA-QSGLASAPVSGAGNHSYCRNPDE 75

Db    447 DAHGPWCY-TGNPLIP-WDYCPISRCEGDTTPTI 478
          |::||||| :|:: :|    |   ||   |: ::
Qy     76 DPRGPWCYVSGEAGVPEKRPCEDLRCPETTSQAL 109


RESULT   11
ID   HGF_HUMAN      STANDARD;      PRT;   728 AA.
AC   P14210;
DT   01-JAN-1990 (REL. 13, CREATED)
DT   01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   HEPATOCYTE GROWTH FACTOR PRECURSOR (SCATTER FACTOR) (SF)
DE   (HEPATOPOEITIN A).
GN   HGF OR HPTA.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91340155.
RA   SEKI T., HAGIYA M., SHIMONISHI M., NAKAMURA T., SHIMIZU S.;
RT   "Organization of the human hepatocyte growth factor-encoding gene.";
RL   GENE 102:213-219(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 89392017.
RA   MIYAZAWA K., TSUBOUCHI H., NAKA D., TAKAHASHI K., OKIGAKI M.,
```

```
RA   ARAKAKI N., NAKAYAMA H., HIRONO S., SAKIYAMA O., TAKAHASHI K.,
RA   GOHDA E., DAIKUHARA Y., KITAMURA N.;
RT   "Molecular cloning and sequence analysis of cDNA for human hepatocyte
RT   growth factor.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 163:967-973(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LEUKOCYTE;
RX   MEDLINE; 91025062.
RA   SEKI T., IHARA I., SUGIMURA A., SHIMONISHI M., NISHIZAWA T.,
RA   ASAMI O., HAGIYA M., NAKAMURA T., SHIMIZU S.;
RT   "Isolation and expression of cDNA for different forms of hepatocyte
RT   growth factor from human leukocyte.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 172:321-327(1990).
RN   [4]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 55-73 AND 495-520.
RC   TISSUE=LIVER;
RX   MEDLINE; 90066676.
RA   NAKAMURA T., NISHIZAWA T., HAGIYA M., SEKI T., SHIMONISHI M.,
RA   SUGIMURA A., TASHIRO K., SHIMIZU S.;
RT   "Molecular cloning and expression of human hepatocyte growth factor.";
RL   NATURE 342:440-443(1989).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYONIC FIBROBLAST;
RX   MEDLINE; 91334393.
RA   WEIDNER K.M., ARAKAKI N., HARTMANN G., VANDEKERCKHOVE J., WEINGART S.,
RA   RIEDER H., FONATSCH C., TSUBOUCHI H., HISHIDA T., DAIKUHARA Y.,
RA   BIRCHMEIER W.;
RT   "Evidence for the identity of human scatter factor and human
RT   hepatocyte growth factor.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 88:7001-7005(1991).
RN   [6]
RP   SIGNAL SEQUENCE CLEAVAGE SITE.
RX   MEDLINE; 91207365.
RA   YOSHIYAMA Y., ARAKAKAI N., NAKA D., TAKAHASHI K., HIRONO S., KONDO J.,
RA   NAKAYAMA H., GOHDA E., KITAMURA N., TSUBOUCHI H., ISHII T.,
RA   HISHIDA T., DAIKUHARA Y.;
RT   "Identification of the N-terminal residue of the heavy chain of both
RT   native and recombinant human hepatocyte growth factor.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 175:660-667(1991).
RN   [7]
RP   CARBOHYDRATE-BINDING SITE 476.
RX   MEDLINE; 93129192.
RA   SHIMIZU N., HARA H., SOGABE T., SAKAI H., IHARA I., INOUE H.,
RA   NAKAMURA T., SHIMIZU S.;
RT   "Hepatocyte growth factor is linked by O-glycosylated oligosaccharide
RT   on the alpha chain.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 189:1329-1335(1992).
RN   [8]
RP   MUTAGENESIS.
RX   MEDLINE; 92331602.
RA   LOKKER N.A., MARK M.R., LUIS E.A., BENNETT G.L., ROBBINS K.A.,
RA   BAKER J.B., GODOWSKI P.J.;
RT   "Structure-function analysis of hepatocyte growth factor:
RT   identification of variants that lack mitogenic activity yet retain
RT   high affinity receptor binding.";
RL   EMBO J. 11:2503-2510(1992).
RN   [9]
RP   STRUCTURE BY NMR OF 31-127.
RX   MEDLINE; 98154323.
RA   ZHOU H., MAZZULLA M.J., KAUFMAN J.D., STAHL S.J., WINGFIELD P.T.,
RA   RUBIN J.S., BOTTARO D.P., BYRD R.A.;
RT   "The solution structure of the N-terminal domain of hepatocyte growth
RT   factor reveals a potential heparin-binding site.";
RL   STRUCTURE 6:109-116(1998).
RN   [10]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 35-210.
RX   MEDLINE; 99036858.
RA   ULTSCH M., LOKKER N.A., GODOWSKI P.J., DE VOS A.M.;
RT   "Crystal structure of the NK1 fragment of human hepatocyte growth
RT   factor at 2.0-A resolution.";
RL   STRUCTURE 6:1383-1393(1998).
CC   -!- FUNCTION: HGF IS A POTENT MITOGEN FOR MATURE PARENCHYMAL
CC       HEPATOCYTE CELLS, SEEMS TO BE AN HEPATOTROPHIC FACTOR, AND ACTS
CC       AS GROWTH FACTOR FOR A BROAD SPECTRUM OF TISSUES AND CELL TYPES.
CC       IT HAS NO DETECTABLE PROTEASE ACTIVITY.
CC   -!- SUBUNIT: DIMER OF AN ALPHA CHAIN AND A BETA CHAIN LINKED BY A
CC       DISULFIDE BOND.
CC   -!- SIMILARITY: CONTAINS 4 KRINGLE REGIONS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D90334; G219700; -.
DR   EMBL; D90318; G219700; JOINED.
DR   EMBL; D90319; G219700; JOINED.
DR   EMBL; D90320; G219700; JOINED.
DR   EMBL; D90322; G219700; JOINED.
DR   EMBL; D90323; G219700; JOINED.
DR   EMBL; D90324; G219700; JOINED.
DR   EMBL; D90325; G219700; JOINED.
DR   EMBL; D90326; G219700; JOINED.
DR   EMBL; D90327; G219700; JOINED.
DR   EMBL; D90328; G219700; JOINED.
DR   EMBL; D90329; G219700; JOINED.
DR   EMBL; D90330; G219700; JOINED.
DR   EMBL; D90331; G219700; JOINED.
DR   EMBL; D90332; G219700; JOINED.
DR   EMBL; D90333; G219700; JOINED.
DR   EMBL; M29145; G306846; -.
DR   EMBL; M60718; G184032; -.
DR   EMBL; X16323; G32082; -.
DR   EMBL; M73239; G337936; -.
DR   EMBL; M73240; G337938; -.
DR   PIR; JH0579; JH0579.
DR   PIR; S06794; S06794.
DR   PDB; 2HGF; 24-JUN-98.
DR   PDB; 1BHT; 18-NOV-98.
DR   MIM; 142409; -.
DR   PROSITE; PS00021; KRINGLE_1; 4.
DR   PROSITE; PS50070; KRINGLE_2; 4.
DR   PFAM; PF00051; kringle; 4.
DR   PFAM; PF00089; trypsin; 1.
KW   GROWTH FACTOR; KRINGLE; GLYCOPROTEIN; SERINE PROTEASE HOMOLOG;
KW   SIGNAL; 3D-STRUCTURE.
FT   SIGNAL       1    31
FT   CHAIN       32   494       HEPATOCYTE GROWTH FACTOR ALPHA CHAIN.
FT   CHAIN      495   728       HEPATOCYTE GROWTH FACTOR BETA CHAIN.
FT   MOD_RES     32    32       PYRROLIDONE CARBOXYLIC ACID.
FT   DOMAIN      32   127       PAP.
FT   DOMAIN     128   206       KRINGLE 1.
FT   DOMAIN     211   288       KRINGLE 2.
FT   DOMAIN     305   383       KRINGLE 3.
FT   DOMAIN     391   469       KRINGLE 4.
FT   DOMAIN     495   728       SERINE PROTEASE-LIKE.
FT   DISULFID    70    96
FT   DISULFID    74    84
FT   DISULFID   128   206
FT   DISULFID   149   189
FT   DISULFID   177   201
FT   DISULFID   487   604       INTERCHAIN (BY SIMILARITY).
FT   CARBOHYD   294   294       POTENTIAL.
FT   CARBOHYD   402   402       POTENTIAL.
FT   CARBOHYD   476   476
FT   CARBOHYD   566   566       POTENTIAL.
FT   CARBOHYD   653   653       POTENTIAL.
FT   CONFLICT    32    33       QR -> HK (IN REF. 4).
FT   CONFLICT    78    78       K -> N (IN REF. 4).
```

```
FT   CONFLICT    293    293       M -> V (IN REF. 4).
FT   CONFLICT    300    300       L -> M (IN REF. 4).
FT   CONFLICT    317    317       V -> A (IN REF. 4).
FT   CONFLICT    336    336       E -> K (IN REF. 4).
FT   CONFLICT    387    387       H -> N (IN REF. 4).
FT   CONFLICT    416    416       D -> N (IN REF. 4).
FT   CONFLICT    505    505       I -> V (IN REF. 4).
FT   CONFLICT    509    509       V -> I (IN REF. 4).
FT   CONFLICT    558    558       D -> E (IN REF. 4).
FT   CONFLICT    561    561       C -> R (IN REF. 4).
FT   CONFLICT    595    595       S -> N (IN REF. 4).
SQ   SEQUENCE   728 AA;  83133 MW;  5182013A CRC32;

  Query Match             8.9%;  Score 167;  DB 1;  Length 728;
  Best Local Similarity  31.5%;  Pred. No. 2.53e-13;
  Matches    28;  Conservative   22;  Mismatches  33;  Indels   6;  Gaps   4;

Db    391 CYRGNGKNYMGNLSQTRSGLTCSMWDKNMEDLHRHIFWEPDASKLNENYCRNPDDDAHGP 450
          |:: || |  : :   :|| |  |   ::  :  :   | ::  | :||||||:|::||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW---LDA-QSGLASAPVSGAGNHSYCRNPDEDPRGP 80

Db    451 WCY-TGNPLIP-WDYCPISRCEGDTTPTI 477
          ||| :|:: :|    |   ||   |: ::
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQAL 109


RESULT   12
ID   THRB_HUMAN     STANDARD;      PRT;   622 AA.
AC   P00734;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   PROTHROMBIN PRECURSOR (EC 3.4.21.5) (COAGULATION FACTOR II).
GN   F2.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88077877.
RA   DEGEN S.J.F., DAVIE E.W.;
RT   "Nucleotide sequence of the gene for human prothrombin.";
RL   BIOCHEMISTRY 26:6165-6177(1987).
RN   [2]
RP   SEQUENCE OF 8-622 FROM N.A.
RX   MEDLINE; 83231469.
RA   DEGEN S.J.F., MCGILLIVRAY R.T.A., DAVIE E.W.;
RT   "Characterization of the complementary deoxyribonucleic acid and gene
RT   coding for human prothrombin.";
RL   BIOCHEMISTRY 22:2087-2097(1983).
RN   [3]
RP   SEQUENCE OF 44-314.
RX   MEDLINE; 77193964.
RA   WALZ D.A., HEWETT-EMMETT D., SEEGERS W.H.;
RT   "Amino acid sequence of human prothrombin fragments 1 and 2.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 74:1969-1972(1977).
RN   [4]
RP   SEQUENCE OF 315-622.
RX   MEDLINE; 77207112.
RA   BUTKOWSKI R.J., ELION J., DOWNING M.R., MANN K.G.;
RT   "Primary structure of human prethrombin 2 and alpha-thrombin.";
RL   J. BIOL. CHEM. 252:4942-4957(1977).
RN   [5]
RP   PROCESSING.
RX   MEDLINE; 87008532.
RA   RABIET M.J., BLASHILL A., FURIE B., FURIE B.C.;
RT   "Prothrombin fragment 1 X 2 X 3, a major product of prothrombin
RT   activation in human plasma.";
RL   J. BIOL. CHEM. 261:13210-13215(1986).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX   MEDLINE; 90059942.
RA   BODE W., MAYR I., BAUMANN U., HUBER R., STONE S.R., HOFSTEENGE J.;
RT   "The refined 1.9 A crystal structure of human alpha-thrombin:
RT   interaction with D-Phe-Pro-Arg chloromethylketone and significance of
RT   the Tyr-Pro-Pro-Trp insertion segment.";
RL   EMBO J. 8:3467-3475(1989).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE; 90327074.
RA   RYDEL T.J., RAVICHANDRAN K.G., TULINSKY A., BODE W., HUBER R.,
RA   ROITSCH C., FENTON J.W. II;
RT   "The structure of a complex of recombinant hirudin and human alpha-
RT   thrombin.";
RL   SCIENCE 249:277-280(1990).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX   MEDLINE; 94350942.
RA   RYDEL T.J., YIN M., PADMANABHAN K.P., BLANKENSHIP D.T., CARDIN A.D.,
RA   CORREA P.E., FENTON J.W. II, TULINSKY A.;
RT   "Crystallographic structure of human gamma-thrombin.";
RL   J. BIOL. CHEM. 269:22000-22006(1994).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE; 97357286.
RA   VAN DE LOCHT A., BODE W., HUBER R., LE BONNIEC B.F., STONE S.R.,
RA   ESMON C.T., STUBBS M.T.;
RT   "The thrombin E192Q-BPTI complex reveals gross structural
RT   rearrangements: implications for the interaction with antithrombin
RT   and thrombomodulin.";
RL   EMBO J. 16:2977-2984(1997).
RN   [10]
RP   VARIANT BARCELONA.
RX   MEDLINE; 87033739.
RA   RABIET M.-J., FURIE B.C., FURIE B.;
RT   "Molecular defect of prothrombin Barcelona. Substitution of cysteine
RT   for arginine at residue 273.";
RL   J. BIOL. CHEM. 261:15045-15048(1986).
RN   [11]
RP   VARIANT FRANKFURT.
RX   MEDLINE; 95313001.
RA   DEGEN S.J.F., MCDOWELL S.A., SPARKS L.M., SCHARRER I.;
RT   "Prothrombin Frankfurt: a dysfunctional prothrombin characterized by
RT   substitution of Glu-466 by Ala.";
RL   THROMB. HAEMOST. 73:203-209(1995).
RN   [12]
RP   VARIANTS HIMI-1 AND HIMI-2.
RX   MEDLINE; 93043342.
RA   MORISHITA E., SAITO M., KUMABASHIRI I., ASAKURA H., MATSUDA T.,
RA   YAMAGUCHI K.;
RT   "Prothrombin Himi: a compound heterozygote for two dysfunctional
RT   prothrombin molecules (Met-337-->Thr and Arg-388-->His).";
RL   BLOOD 80:2275-2280(1992).
RN   [13]
RP   VARIANT PADUA-1.
RX   MEDLINE; 95169898.
RA   JAMES H.L., KIM D.J., ZHENG D.-Q., GIROLAMI A.;
RT   "Prothrombin Padua I: incomplete activation due to an amino acid
RT   substitution at a factor Xa cleavage site.";
RL   BLOOD COAGUL. FIBRINOLYSIS 5:841-844(1994).
RN   [14]
RP   VARIANT QUICK-1.
RX   MEDLINE; 89207504.
RA   HENRIKSEN R.A., MANN K.G.;
RT   "Identification of the primary structural defect in the dysthrombin
RT   thrombin Quick I: substitution of cysteine for arginine-382.";
RL   BIOCHEMISTRY 27:9160-9165(1988).
RN   [15]
RP   VARIANT QUICK-2.
RX   MEDLINE; 89247398.
RA   HENRIKSEN R.A., MANN K.G.;
RT   "Substitution of valine for glycine-558 in the congenital dysthrombin
RT   thrombin Quick II alters primary substrate specificity.";
RL   BIOCHEMISTRY 28:2078-2082(1989).
RN   [16]
RP   VARIANT SALAKTA.
```

```
RX   MEDLINE; 92378975.
RA   MIYATA T., ARUGA R., UMEYAMA H., BEZEAUD A., GUILLIN M.-C.,
RA   IWANAGA S.;
RT   "Prothrombin Salakta: substitution of glutamic acid-466 by alanine
RT   reduces the fibrinogen clotting activity and the esterase activity.";
RL   BIOCHEMISTRY 31:7457-7462(1992).
RN   [17]
RP   VARIANT TOKUSHIMA.
RX   MEDLINE; 87185407.
RA   MIYATA T., MORITA T., INOMOTO T., KAWAUCHI S., SHIRAKAMI A.,
RA   IWANAGA S.;
RT   "Prothrombin Tokushima, a replacement of arginine-418 by tryptophan
RT   that impairs the fibrinogen clotting activity of derived thrombin
RT   Tokushima.";
RL   BIOCHEMISTRY 26:1117-1122(1987).
RN   [18]
RP   VARIANT TOKUSHIMA.
RX   MEDLINE; 87101511.
RA   INOMOTO T., SHIRAKAMI A., KAWAUCHI S., SHIGEKIYO T., SAITO S.,
RA   MIYOSHI K., MORITA T., IWANAGA S.;
RT   "Prothrombin Tokushima: characterization of dysfunctional thrombin
RT   derived from a variant of human prothrombin.";
RL   BLOOD 69:565-569(1987).
RN   [19]
RP   VARIANT TOKUSHIMA.
RX   MEDLINE; 92256895.
RA   IWAHANA H., YOSHIMOTO K., SHIGEKIYO T., SHIRAKAMI A., SAITO S.,
RA   ITAKURA M.;
RT   "Detection of a single base substitution of the gene for prothrombin
RT   Tokushima. The application of PCR-SSCP for the genetic and molecular
RT   analysis of dysprothrombinemia.";
RL   INT. J. HEMATOL. 55:93-100(1992).
RN   [20]
RP   VARIANT TYPE-3.
RX   MEDLINE; 83204687.
RA   BOARD P.G., SHAW D.C.;
RT   "Determination of the amino acid substitution in human prothrombin
RT   type 3 (157 Glu leads to Lys) and the localization of a third
RT   thrombin cleavage site.";
RL   BR. J. HAEMATOL. 54:245-254(1983).
CC   -!- FUNCTION: THROMBIN, WHICH CLEAVES BONDS AFTER ARG & LYS, CONVERTS
CC       FIBRINOGEN TO FIBRIN AND ACTIVATES FACTORS V, VII, VIII, XIII,
CC       AND, IN COMPLEX WITH THROMBOMODULIN, PROTEIN C.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: SYNTHESIZED IN THE LIVER; FOUND IN PLASMA.
CC   -!- PTM: THE GAMMA-CARBOXYGLUTAMYL RESIDUES, WHICH BIND CALCIUM IONS,
CC       RESULT FROM THE CARBOXYLATION OF GLUTAMYL RESIDUES BY A MICROSOMAL
CC       ENZYME, THE VITAMIN K-DEPENDENT CARBOXYLASE. THE MODIFIED RESIDUES
CC       ARE NECESSARY FOR THE CA-DEPENDENT INTERACTION WITH A NEGATIVELY
CC       CHARGED PHOSPHOLIPID SURFACE,WHICH IS ESSENTIAL FOR THE CONVERSION
CC       OF PROTHROMBIN TO THROMBIN.
CC   -!- DISEASE: DEFECTS IN F2 ARE THE CAUSE OF VARIOUS FORMS OF
CC       DYSPROTHROMBINEMIA.
CC   -!- PROTHROMBIN IS ACTIVATED ON THE SURFACE OF A PHOSPHOLIPID
CC       MEMBRANE THAT BINDS THE AMINO END OF PROTHROMBIN & FACTORS VA & XA
CC       IN CA-DEPENDENT INTERACTIONS; FACTOR XA REMOVES THE ACTIVATION
CC       PEPTIDE & CLEAVES THE REMAINING PART INTO LIGHT & HEAVY CHAINS.
CC       THE ACTIVATION PROCESS STARTS SLOWLY BECAUSE FACTOR V ITSELF HAS
CC       TO BE ACTIVATED BY THE INITIAL, SMALL AMOUNTS OF THROMBIN.
CC   -!- IT IS NOT KNOWN WHETHER 1 OR 2 SMALLER ACTIVATION PEPTIDES, WITH
CC       ADDITIONAL CLEAVAGE AFTER 314-ARG, ARE RELEASED IN NATURAL BLOOD
CC       CLOTTING.
CC   -!- THROMBIN CAN ITSELF CLEAVE THE AMINO TERMINAL FRAGMENT (FRAGMENT
CC       1) OF THE PROTHROMBIN, PRIOR TO ITS ACTIVATION BY FACTOR XA.
CC   -!- THE CLEAVAGE AFTER R-198, OBSERVED IN VITRO, DOES NOT OCCUR IN
CC       PLASMA.
CC   -!- SIMILARITY: CONTAINS 2 KRINGLE REGIONS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M17262; G339641; -.
DR   EMBL; V00595; E5121; -.
DR   PIR; A00914; TBHU.
...
Note: remainder of annotations omitted.

  Query Match             8.8%;  Score 165;  DB 1;  Length 622;
  Best Local Similarity  37.7%;  Pred. No. 6.12e-13;
  Matches    26;  Conservative   13;  Mismatches  26;  Indels   4;  Gaps   3;

Db    213 CVPDRGQQYQGRLAVTTHGLPCLAWASAQAKALSKHQDFNSAVQLVENFCRNPDGDEEGV 272
          |  | |: |:   :  : || || |  ||: :|:     ::|     ::||||| |  |
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLDAQS-GLASAP-VSGAGN--HSYCRNPDEDPRGP 80

Db    273 WCYVAGKPG 281
          ||||:| :|
Qy     81 WCYVSGEAG 89


RESULT   13
ID   HGFL_HUMAN     STANDARD;      PRT;   711 AA.
AC   P26927;
DT   01-AUG-1992 (REL. 23, CREATED)
DT   01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   HEPATOCYTE GROWTH FACTOR-LIKE PROTEIN PRECURSOR (MACROPHAGE
DE   STIMULATORY PROTEIN) (MSP) (MACROPHAGE STIMULATING PROTEIN).
GN   MST1 OR HGFL.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 92002016.
RA   HAN S., STUART L.A., FRIEZNER DEGEN S.J.;
RT   "Characterization of the DNF15S2 locus on human chromosome 3:
RT   identification of a gene coding for four kringle domains with
RT   homology to hepatocyte growth factor.";
RL   BIOCHEMISTRY 30:9768-9780(1991).
CC   -!- FUNCTION: PROBABLY HAS NO PROTEOLYTIC ACTIVITY, SINCE CRUCIAL AA
CC       CHARACTERISTIC OF SERINE PROTEASES CATALYTIC SITES ARE NOT
CC       CONSERVED.
CC   -!- PTM: MAY BE CLEAVED AFTER AA 484, TO YIELD A TWO-CHAIN MOLECULE
CC       HELD TOGETHER BY DISULFIDE BONDS, OR TWO SEPARATE POLYPEPTIDES.
CC   -!- SIMILARITY: CONTAINS 4 KRINGLE REGIONS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M74178; G183977; -.
DR   EMBL; M74179; G183979; -.
DR   PIR; A40331; A40331.
DR   MIM; 142408; -.
DR   PROSITE; PS00021; KRINGLE_1; 4.
DR   PROSITE; PS50070; KRINGLE_2; 4.
DR   PFAM; PF00051; kringle; 4.
DR   PFAM; PF00089; trypsin; 1.
DR   HSSP; P00763; 1SLW.
KW   KRINGLE; GLYCOPROTEIN; SERINE PROTEASE HOMOLOG; SIGNAL;
KW   POLYMORPHISM.
```

```
FT   SIGNAL        1     31       POTENTIAL.
FT   CHAIN        32    711       HEPATOCYTE GROWTH FACTOR-LIKE PROTEIN.
FT   DOMAIN       32    109       PAP.
FT   DOMAIN      110    186       KRINGLE 1.
FT   DOMAIN      191    268       KRINGLE 2.
FT   DOMAIN      283    361       KRINGLE 3.
FT   DOMAIN      370    448       KRINGLE 4.
FT   DOMAIN      484    711       SERINE PROTEASE-LIKE.
FT   DISULFID     56     78       BY SIMILARITY.
FT   DISULFID     60     66       BY SIMILARITY.
FT   DISULFID    110    186       BY SIMILARITY.
FT   DISULFID    131    169       BY SIMILARITY.
FT   DISULFID    157    181       BY SIMILARITY.
FT   DISULFID    191    268       BY SIMILARITY.
FT   DISULFID    194    324       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    212    251       BY SIMILARITY.
FT   DISULFID    240    263       BY SIMILARITY.
FT   DISULFID    283    361       BY SIMILARITY.
FT   DISULFID    304    343       BY SIMILARITY.
FT   DISULFID    332    355       BY SIMILARITY.
FT   DISULFID    370    448       BY SIMILARITY.
FT   DISULFID    391    431       BY SIMILARITY.
FT   DISULFID    419    443       BY SIMILARITY.
FT   DISULFID    468    588       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    507    523       BY SIMILARITY.
FT   DISULFID    602    667       BY SIMILARITY.
FT   DISULFID    632    646       BY SIMILARITY.
FT   DISULFID    657    685       BY SIMILARITY.
FT   CARBOHYD     72     72       POTENTIAL.
FT   CARBOHYD    296    296       POTENTIAL.
FT   CARBOHYD    615    615       POTENTIAL.
FT   VARIANT      13     13       Y -> C.
FT   VARIANT     212    212       C -> F.
SQ   SEQUENCE   711 AA;  80379 MW;  C9978F05 CRC32;

  Query Match             8.8%;  Score 166;  DB 1;  Length 711;
  Best Local Similarity  32.5%;  Pred. No. 3.94e-13;
  Matches    27;  Conservative   17;  Mismatches  35;  Indels   4;  Gaps   4;

Db    110 CIMNNGVGYRGTMATTVGGLPCQAWSHKFPNDHKYTPTLRNGLEENFCRNPDGDPGGPWC 169
          |: :||  ||   ::   || |  |     :    :|    | : ::||||| || ||||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLDA-QSGLASAPVSGAG-NHSYCRNPDEDPRGPWC 82

Db    170 YTT-DPAVRF-QSCGIKSCREAA 190
          | : :::|   ::|    | |::
Qy     83 YVSGEAGVPEKRPCEDLRCPETT 105


RESULT   14
ID   HGF_RAT        STANDARD;      PRT;   728 AA.
AC   P17945;
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   HEPATOCYTE GROWTH FACTOR PRECURSOR (SCATTER FACTOR) (SF)
DE   (HEPATOPOEITIN A).
GN   HGF.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=WISTAR; TISSUE=LIVER;
RX   MEDLINE; 90222197.
RA   TOSHIRO K., HAGIYA M., NISHIZAWA T., SEKI T., SHIMONISHI M.,
RA   SHIMIZU S., NAKAMURA T.;
RT   "Deduced primary structure of rat hepatocyte growth factor and
RT   expression of the mRNA in rat tissues.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 87:3200-3204(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR; TISSUE=LIVER;
RX   MEDLINE; 91031482.
RA   OKAJIMA A., MIYAZAWA K., KITAMURA N.;
RT   "Primary structure of rat hepatocyte growth factor and induction of
RT   its mRNA during liver regeneration following hepatic injury.";
RL   EUR. J. BIOCHEM. 193:375-381(1990).
CC   -!- FUNCTION: HGF IS A POTENT MITOGEN FOR MATURE PARENCHYMAL
CC       HEPATOCYTE CELLS, SEEMS TO BE AN HEPATOTROPHIC FACTOR, AND ACTS
CC       AS GROWTH FACTOR FOR A BROAD SPECTRUM OF TISSUES AND CELL TYPES.
CC       IT HAS NO DETECTABLE PROTEASE ACTIVITY.
CC   -!- SUBUNIT: DIMER OF AN ALPHA CHAIN AND A BETA CHAIN LINKED BY A
CC       DISULFIDE BOND.
CC   -!- SIMILARITY: CONTAINS 4 KRINGLE REGIONS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D90102; G220767; -.
DR   EMBL; X54400; G56354; -.
DR   PIR; S13211; S13211.
DR   PIR; A35644; A35644.
DR   PROSITE; PS00021; KRINGLE_1; 4.
DR   PROSITE; PS50070; KRINGLE_2; 4.
DR   PFAM; PF00051; kringle; 4.
DR   PFAM; PF00089; trypsin; 1.
DR   HSSP; P14210; 2HGF.
KW   GROWTH FACTOR; KRINGLE; GLYCOPROTEIN; SERINE PROTEASE HOMOLOG;
KW   SIGNAL.
FT   SIGNAL        1     32       BY SIMILARITY.
FT   CHAIN        33    495       HEPATOCYTE GROWTH FACTOR ALPHA CHAIN.
FT   CHAIN       496    728       HEPATOCYTE GROWTH FACTOR BETA CHAIN.
FT   MOD_RES      33     33       PYRROLIDONE CARBOXYLIC ACID
FT                                (BY SIMILARITY).
FT   DOMAIN       33    128       PAP.
FT   DOMAIN      129    207       KRINGLE 1.
FT   DOMAIN      212    289       KRINGLE 2.
FT   DOMAIN      306    384       KRINGLE 3.
FT   DOMAIN      392    470       KRINGLE 4.
FT   DOMAIN      496    728       SERINE PROTEASE-LIKE.
FT   DISULFID     71     97       BY SIMILARITY.
FT   DISULFID     75     85       BY SIMILARITY.
FT   DISULFID    488    607       INTERCHAIN (BY SIMILARITY).
FT   CARBOHYD    295    295       POTENTIAL.
FT   CARBOHYD    403    403       POTENTIAL.
FT   CARBOHYD    569    569       POTENTIAL.
FT   CARBOHYD    656    656       POTENTIAL.
SQ   SEQUENCE   728 AA;  82905 MW;  0285098A CRC32;

  Query Match             8.8%;  Score 165;  DB 1;  Length 728;
  Best Local Similarity  31.9%;  Pred. No. 6.12e-13;
  Matches    30;  Conservative   22;  Mismatches  35;  Indels   7;  Gaps   5;

Db    387 SSGQDCYRGNGKNYMGNLSKTRSGLTCSMWDKNMEDLHRHIFWEPDASKLTKNYCRNPDD 446
          :||  |:: ||  |  : :   :|| |  |   ::  :  :   | ::    :||||||:
Qy     21 GSG-GCFWDNGHLYREDQTSPAPGLRCLNW---LDA-QSGLASAPVSGAGNHSYCRNPDE 75

Db    447 DAHGPWCY-TGNPLVP-WDYCPISRCEGDTTPTI 478
          |::||||| :|:: ||    |   ||   |: ::
Qy     76 DPRGPWCYVSGEAGVPEKRPCEDLRCPETTSQAL 109


RESULT   15
ID   THRB_BOVIN     STANDARD;      PRT;   625 AA.
AC   P00735;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   PROTHROMBIN PRECURSOR (EC 3.4.21.5).
```

```
GN   F2.
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88245190.
RA   IRWIN D.M., ROBERTSON K.A., MACGILLIVRAY R.T.A.;
RT   "Structure and evolution of the bovine prothrombin gene.";
RL   J. MOL. BIOL. 200:31-45(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 84203525.
RA   MCGILLIVRAY R.T.A., DAVIE E.W.;
RT   "Characterization of bovine prothrombin mRNA and its translation
RT   product.";
RL   BIOCHEMISTRY 23:1626-1634(1984).
RN   [3]
RP   SEQUENCE OF 44-625, DISULFIDE BONDS, AND CARBOHYDRATE-BINDING SITES.
RA   MAGNUSSON S., SOTTRUP-JENSEN L., PETERSEN T.E., CLAEYS H.;
RL   (IN) HEMKER H.C., VELTKAMP J.J. (EDS.);
RL   BOERHAAVE SYMPOSIUM ON PROTHROMBIN AND RELATED COAGULATION FACTORS,
RL   PP.25-46, LEIDEN UNIVERSITY PRESS, LEIDEN (1975).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF ACTIVATION PEPTIDE 1.
RX   MEDLINE; 86296631.
RA   PARK C.H., TULINSKY A.;
RT   "Three-dimensional structure of the kringle sequence: structure of
RT   prothrombin fragment 1.";
RL   BIOCHEMISTRY 25:3977-3982(1986).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS) OF ACTIVATION PEPTIDE 1.
RX   MEDLINE; 91311686.
RA   SESHADRI T.-P., TULINSKY A., SKRZYPCZAK-JANKUN E., PARK C.H.;
RT   "Structure of bovine prothrombin fragment 1 refined at 2.25-A
RT   resolution.";
RL   J. MOL. BIOL. 220:481-494(1991).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF ACTIVATION PEPTIDE 1.
RX   MEDLINE; 92190185.
RA   SORIANO-GARCIA M., PADMANABHAN K., DE VOS A.M., TULINSKY A.;
RT   "The Ca2+ ion and membrane binding structure of the Gla domain of Ca-
RT   prothrombin fragment 1.";
RL   BIOCHEMISTRY 31:2554-2566(1992).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE; 92218459.
RA   MARTIN P.D., ROBERTSON W., TURK D., HUBER R., BODE W., EDWARDS B.F.P.;
RT   "The structure of residues 7-16 of the A alpha-chain of human
RT   fibrinogen bound to bovine thrombin at 2.3-A resolution.";
RL   J. BIOL. CHEM. 267:7911-7920(1992).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE; 92389319.
RA   BRANDSTETTER H., TURK D., HOEFFKEN H.W., GROSSE D., STUERZEBECHER J.,
RA   MARTIN P.D., EDWARDS B.F.P., BODE W.;
RT   "Refined 2.3 A X-ray crystal structure of bovine thrombin complexes
RT   formed with the benzamidine and arginine-based thrombin inhibitors
RT   NAPAP, 4-TAPAP and MQPA. A starting point for improving
RT   antithrombotics.";
RL   J. MOL. BIOL. 226:1085-1089(1992).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF COMPLEX WITH ORNITHODORIN.
RX   MEDLINE; 97102783.
RA   VAN DE LOCHT A., STUBBS M.T., BODE W., FRIEDRICH T., BOLLSCHWEILER C.,
RA   HOFFKEN W., HUBER R.;
RT   "The ornithodorin-thrombin crystal structure, a key to the TAP
RT   enigma?";
RL   EMBO J. 15:6011-6017(1996).
RN   [10]
RP   GENE STRUCTURE.
RX   MEDLINE; 86077733.
RA   IRWIN D.M., AHERN K.G., PEARSON G.D., MCGILLIVRAY R.T.A.;
RT   "Characterization of the bovine prothrombin gene.";
RL   BIOCHEMISTRY 24:6854-6861(1985).
CC   -!- FUNCTION: THROMBIN, WHICH CLEAVES BONDS AFTER ARG & LYS, CONVERTS
CC       FIBRINOGEN TO FIBRIN AND ACTIVATES FACTORS V, VII, VIII, XIII,
CC       AND, IN COMPLEX WITH THROMBOMODULIN, PROTEIN C.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: SYNTHESIZED IN THE LIVER; FOUND IN PLASMA.
CC   -!- PTM: THE GAMMA-CARBOXYGLUTAMYL RESIDUES, WHICH BIND CALCIUM IONS,
CC       RESULT FROM THE CARBOXYLATION OF GLUTAMYL RESIDUES BY A MICROSOMAL
CC       ENZYME, THE VITAMIN K-DEPENDENT CARBOXYLASE. THE MODIFIED RESIDUES
CC       ARE NECESSARY FOR THE CA-DEPENDENT INTERACTION WITH A NEGATIVELY
CC       CHARGED PHOSPHOLIPID SURFACE,WHICH IS ESSENTIAL FOR THE CONVERSION
CC       OF PROTHROMBIN TO THROMBIN.
CC   -!- PROTHROMBIN IS ACTIVATED ON THE SURFACE OF A PHOSPHOLIPID
CC       MEMBRANE THAT BINDS THE AMINO END OF PROTHROMBIN & FACTORS VA & XA
CC       IN CA-DEPENDENT INTERACTIONS; FACTOR XA REMOVES THE ACTIVATION
CC       PEPTIDE & CLEAVES THE REMAINING PART INTO LIGHT & HEAVY CHAINS.
CC       THE ACTIVATION PROCESS STARTS SLOWLY BECAUSE FACTOR V ITSELF HAS
CC       TO BE ACTIVATED BY THE INITIAL, SMALL AMOUNTS OF THROMBIN.
CC   -!- THROMBIN CAN ITSELF CLEAVE THE AMINO TERMINAL FRAGMENT (FRAGMENT
CC       1) OF THE PROTHROMBIN, PRIOR TO ITS ACTIVATION BY FACTOR XA.
CC   -!- SIMILARITY: CONTAINS 2 KRINGLE REGIONS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; V00135; G808945; -.
DR   EMBL; J00041; G163755; -.
DR   PIR; A00915; TBBO.
DR   PIR; S02537; S02537.
DR   PDB; 1BBR; 31-JAN-94.
DR   PDB; 1ETR; 31-JAN-94.
DR   PDB; 1ETS; 31-JAN-94.
DR   PDB; 1ETT; 31-JAN-94.
DR   PDB; 1HRT; 31-JAN-94.
DR   PDB; 2PF1; 31-JAN-94.
DR   PDB; 2PF2; 31-JAN-94.
DR   PDB; 2SPT; 31-MAY-94.
DR   PDB; 1MKW; 07-JUL-97.
DR   PDB; 1MKX; 07-JUL-97.
DR   PDB; 1TBQ; 14-OCT-96.
DR   PDB; 1TBR; 14-OCT-96.
DR   PDB; 1TOC; 23-JUL-97.
DR   PDB; 1VIT; 21-APR-97.
DR   PDB; 1YCP; 06-MAY-98.
DR   PDB; 1A0H; 17-JUN-98.
DR   PROSITE; PS00011; GLU_CARBOXYLATION; 1.
DR   PROSITE; PS00021; KRINGLE_1; 2.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PROSITE; PS50070; KRINGLE_2; 2.
DR   PFAM; PF00051; kringle; 2.
DR   PFAM; PF00089; trypsin; 1.
DR   PFAM; PF00594; gla; 1.
KW   BLOOD COAGULATION; PLASMA; CALCIUM-BINDING; GLYCOPROTEIN; DUPLICATION;
KW   VITAMIN K; ZYMOGEN; GAMMA-CARBOXYGLUTAMIC ACID; ACUTE PHASE; LIVER;
KW   HYDROLASE; SERINE PROTEASE; KRINGLE; SIGNAL; 3D-STRUCTURE.
FT   SIGNAL       1     ?
FT   PROPEP       ?     43
FT   CHAIN        44    625       PROTHROMBIN.
FT   PEPTIDE      44    199       ACTIVATION PEPTIDE (FRAGMENT 1).
FT   PEPTIDE      200   317       ACTIVATION PEPTIDE (FRAGMENT 2).
FT   CHAIN        318   366       THROMBIN LIGHT CHAIN (A).
FT   CHAIN        367   625       THROMBIN HEAVY CHAIN (B).
FT   DOMAIN       109   187       KRINGLE 1.
FT   DOMAIN       214   292       KRINGLE 2.
```

```
FT   SITE        199    200       CLEAVAGE (BY THROMBIN).
FT   SITE        317    318       CLEAVAGE (BY FACTOR XA).
FT   SITE        366    367       CLEAVAGE (BY FACTOR XA).
FT   ACT_SITE    409    409       CHARGE RELAY SYSTEM.
FT   ACT_SITE    465    465       CHARGE RELAY SYSTEM.
FT   ACT_SITE    571    571       CHARGE RELAY SYSTEM.
FT   MOD_RES      50     50       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      51     51       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      58     58       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      60     60       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      63     63       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      64     64       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      69     69       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      70     70       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      73     73       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      76     76       GAMMA-CARBOXYGLUTAMIC ACID.
FT   CARBOHYD    120    120
FT   CARBOHYD    144    144
FT   CARBOHYD    419    419
FT   DISULFID     61     66
FT   DISULFID     91    104
FT   DISULFID    109    187
FT   DISULFID    130    170
FT   DISULFID    158    182
FT   DISULFID    214    292
FT   DISULFID    235    275
FT   DISULFID    263    287
FT   DISULFID    339    485       INTERCHAIN.
FT   DISULFID    394    410
FT   DISULFID    539    553
FT   DISULFID    567    597
FT   VARIANT     600    600       D -> N.
FT   CONFLICT    231    231       S -> H (IN REF. 2).
FT   CONFLICT    249    249       D -> H (IN REF. 2).
FT   CONFLICT    288    288       D -> N (IN REF. 3).
FT   CONFLICT    353    353       Q -> E (IN REF. 3).
FT   CONFLICT    355    355       E -> Q (IN REF. 3).
FT   CONFLICT    549    550       DN -> ND (IN REF. 3).
FT   HELIX        50     52
FT   HELIX        57     62
FT   HELIX        68     73
FT   TURN         74     75
FT   HELIX        79     89
FT   TURN         90     93
FT   HELIX        99    105
FT   TURN        106    107
FT   STRAND      110    110
FT   TURN        112    114
...
Note: remainder of annotations omitted.

  Query Match            8.7%;  Score 164;  DB 1;  Length 625;
  Best Local Similarity  36.2%;  Pred. No. 9.51e-13;
  Matches    25;  Conservative   14;  Mismatches  26;  Indels   4;  Gaps   4;

Db    214 CVPDRGREYRGRLAVTTSGSRCLAWSSEQAKALSKDQDFNPAVPLAENFCRNPDGDEEGA 273
          |  | |: ||    :  ::| ||| |   }: :|:     : |     ::||||| |  |:
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLDAQS-GLASAP-VSGA-G-NHSYCRNPDEDPRGP 80

Db    274 WCYVADQPG 282
          |||| : ::|
Qy     81 WCYVSGEAG 89


Search completed: Fri Sep 17 18:27:21 1999
Job time : 25 secs.
```

Best Available Copy

MPSRCH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Sep 17 18:27:37 1999; MasPar time 17.87 Seconds
803.257 Million cell updates/sec

Tabular output not generated.

Title: >US-09-084-491A-2
Description: (1-263) from US09084491A.pep
Perfect Score: 1883
Sequence: 1 MLLAWVQAFLVSNMLLAEAY..........PVDPQEGSTPLMGQAGTPGA 263

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 45.193; Variance 75.818; scale 0.596

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1881	99.9	263	4	O00318	PUTATIVE PROTEIN.	0.00e+00
2	193	10.2	420	13	Q90504	THROMBIN.	2.18e-18
3	184	9.8	704	13	Q90865	HEPATOCYTE GROWTH FACT	1.37e-16
4	178	9.5	716	11	P70521	HEPATOCYTE GROWTH FACT	2.09e-15
5	177	9.4	685	5	Q24488	NEUROTROPHIC RECEPTOR	3.28e-15
6	172	9.1	717	13	P70006	HEPATOCYTE GROWTH FACT	3.10e-14
7	169	9.0	728	11	Q64007	HEPATOCYTE GROWTH FACT	1.18e-13
8	168	8.9	748	11	Q61662	HEPATOCYTE GROWTH FACT	1.84e-13
9	168	8.9	806	6	O18783	PLASMINOGEN.	1.84e-13
10	165	8.8	597	11	O35727	FACTOR XII.	6.95e-13
11	166	8.8	711	4	Q14870	MACROPHAGE-STIMULATING	4.47e-13
12	166	8.8	711	4	Q13350	HEPATOCYTE GROWTH FACT	4.47e-13
13	161	8.6	810	4	Q15146	PLASMINOGEN PRECURSOR.	4.02e-12
14	156	8.3	334	6	O46507	PLASMINOGEN (FRAGMENT)	3.54e-11
15	157	8.3	716	13	Q91691	GROWTH FACTOR LIVERTIN	2.29e-11
16	155	8.2	567	4	Q13208	HEPATOCYTE GROWTH FACT	5.44e-11
17	151	8.0	943	4	Q01974	PROTEIN-TYROSINE KINAS	3.03e-10
18	147	7.8	385	5	Q25101	SERINE PROTEINASE.	1.66e-09
19	146	7.8	411	4	Q15844	UROKINASE-TYPE PLASMIN	2.52e-09
20	146	7.8	431	4	Q16618	UROKINASE (EC 3.4.99.2	2.52e-09
21	141	7.5	454	6	O46506	APOLIPOPROTEIN A (FRAG	2.04e-08
22	141	7.5	607	13	Q91001	THROMBIN.	2.04e-08
23	142	7.5	710	13	Q91402	HEPATOCYTE GROWTH FACT	1.35e-08
24	139	7.4	323	5	P91823	T22A3.6 PROTEIN.	4.67e-08
25	129	6.9	145	6	Q28911	APOLIPOPROTEIN[A (FRAG	2.70e-06
26	130	6.9	202	13	Q90675	TISSUE-TYPE PLASMINOGE	1.81e-06
27	130	6.9	946	13	Q07153	TYROSINE KINASE RECEPT	1.81e-06
28	126	6.7	215	13	O42341	HGF ALPHA-CHAIN (FRAGM	8.89e-06
29	124	6.6	132	4	Q16609	(APOARGC).	1.95e-05
30	124	6.6	560	4	Q14520	HGF ACTIVATOR LIKE PRO	1.95e-05
31	125	6.6	726	13	Q90978	HEPATOCYTE GROWTH FACT	1.32e-05
32	122	6.5	210	4	Q13494	HGF AGONIST/ANTAGONIST	4.25e-05
33	122	6.5	290	4	Q02935	HEPATOCYTE GROWTH FACT	4.25e-05
34	122	6.5	296	4	Q14519	COMPETITIVE HGF ANTAGO	4.25e-05
35	121	6.4	111	6	O77688	PROTHROMBIN PRECURSOR	6.25e-05
36	118	6.3	937	4	Q01973	PROTEIN-TYROSINE KINAS	1.98e-04
37	119	6.3	2869	6	Q28398	APOLIPOPROTEIN (FRAGME	1.35e-04
38	117	6.2	806	5	Q17576	KIN-8 PROTEIN.	2.89e-04
39	115	6.1	211	11	O55027	HEPATOCYTE GROWTH FACT	6.14e-04
40	114	6.1	714	5	O02001	NEUROSPECIFIC RECEPTOR	8.93e-04
41	115	6.1	761	11	O08762	PROTEASE, SERINE, 12 N	6.14e-04
42	111	5.9	412	11	Q63611	TUMOR-ASSOCIATED ANTIG	2.72e-03
43	105	5.6	458	2	P78256	CDP-DIACYLGLYCEROL--SE	2.38e-02
44	102	5.4	1074	5	Q94046	T13F2.3 PROTEIN.	6.84e-02
45	98	5.2	1003	5	Q21977	CODED FOR BY C. ELEGAN	2.70e-01

ALIGNMENTS

not prior

```
RESULT  1
ID   O00318     PRELIMINARY;     PRT;   263 AA.
AC   O00318;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   PUTATIVE PROTEIN.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   DU Z., SCHEET P., HARPER M.;
RL   SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   WATERSTON R.;
RL   SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AC002073; G2078470; -.
SQ   SEQUENCE   263 AA;  28248 MW;  220695EC CRC32;

  Query Match             99.9%;  Score 1881;  DB 4;  Length 263;
  Best Local Similarity   99.6%;  Pred. No. 0.00e+00;
  Matches  262;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Db      1 MLLAWVQAFLVSNMLLAEAYGSGGCFWDNGHLYREDQTSPAPGLRCLNWLDAQSGLASAP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 MLLAWVQAFLVSNMLLAEAYGSGGCFWDNGHLYREDQTSPAPGLRCLNWLDAQSGLASAP 60

Db     61 VSGAGNHSYCRNPDEDPRGPWCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEAS 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 VSGAGNHSYCRNPDEDPRGPWCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEAS 120

Db    121 EGPGADEVQVFAPANALPARSEAAAVQPVIGISQRVRMNSKEKKDLGTLGYVLGITMMVI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121 EGPGADEVQVFAPANALPARSEAAAVQPVIGISQRVRMNSKEKKDLGTLGYVLGITMMVI 180

Db    181 IIAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRITLPLSAFTNPTCEIVDEKTVVVHT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    181 IIAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRITLPLSAFTNPTCEIVDEKTVVVHT 240

Db    241 SQTPVDPQEGTTPLMGQAGTPGA 263
          ||||||||||:||||||||||||
Qy    241 SQTPVDPQEGSTPLMGQAGTPGA 263
```

```
RESULT    2
ID   Q90504      PRELIMINARY;      PRT;   420 AA.
AC   Q90504;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   THROMBIN.
OS   EPTATRETUS STOUTI.
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MYXINI; MYXINIFORMES;
OC   EPTATRETIDAE; EPTATRETUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 94223694.
RA   BANFIELD D.K., IRWIN D.M., WALZ D.A., MACGILLIVRAY R.T.;
RT   "Evolution of prothrombin: isolation and characterization of the
RT   cDNAs encoding chicken and hagfish prothrombin.";
RL   J. MOL. EVOL. 38:177-187(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 92212913.
RA   BANFIELD D.K., MACGILLIVRAY R.T.;
RT   "Partial characterization of vertebrate prothrombin cDNAs:
RT   amplification and sequence analysis of the B chain of thrombin from
RT   nine different species.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 89:2779-2783(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   BANFIELD D.K.;
RL   SUBMITTED (DEC-1991) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; M81393; G290665; -.
DR   PFAM; PF00051; kringle; 1.
DR   PFAM; PF00089; trypsin; 1.
SQ   SEQUENCE   420 AA;  47888 MW;  DB9E82BA CRC32;

  Query Match           10.2%;  Score 193;  DB 13;  Length 420;
  Best Local Similarity  28.8%;  Pred. No. 2.18e-18;
  Matches    38;  Conservative   31;  Mismatches  59;  Indels   4;  Gaps   4;

Db      17 CVRREGRDYRGDLNITWTGKPCLPWRGSYSNFLPSQFTTAGLTSNYCRNPDGDSEGVWCY 76
           | : :|: || |      |  || |  : | : ::  : ||   :|||||| |: | |||
Qy      25 CFWDNGHLYREDQTSPAPGLRCLNWLDAQSGLASAPVSGAG-NHSYCRNPDEDPRGPWCY 83

Db      77 TKGVEGT-DVDYCQ-LNYCES-GDIFEVGTDEVQLSGRSEGAAEKTLFFNPKTFGNGEEE 133
             |  |   :   |: |   |: :: :   | |:| :: : || |  :|  ::::    |
Qy      84 VSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARSEA 143

Db     134 CGKRPMFELQQK 145
           : :|:: : |:
Qy     144 AAVQPVIGISQR 155


RESULT    3
ID   Q90865      PRELIMINARY;      PRT;   704 AA.
AC   Q90865;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   HEPATOCYTE GROWTH FACTOR-LIKE/MACROPHAGE STIMULATING PROTEIN.
GN   HGF1/MSP.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 96029010.
RA   THERY C., SHARPE M.J., BATLEY S.J., STERN C.D., GHERARDI E.;
RT   "Expression of HGF/SF, HGF1/MSP, and c-met suggests new functions
RT   during early chick development.";
RL   DEV. GENET. 17:90-101(1995).
DR   EMBL; X84043; E136589; -.
DR   PFAM; PF00051; kringle; 4.
DR   PFAM; PF00089; trypsin; 1.
SQ   SEQUENCE   704 AA;  79341 MW;  625C9188 CRC32;

  Query Match           9.8%;  Score 184;  DB 13;  Length 704;
  Best Local Similarity  34.1%;  Pred. No. 1.37e-16;
  Matches    28;  Conservative   16;  Mismatches  34;  Indels   4;  Gaps   4;

Db     108 CIVANGTSYRGTRDTTERGLRCQHWQATTPHDHRFLPSLRNGLEENYCRNPDRDKRGPWC 167
           |:  ||  ||  : :   |||| :|  : :      |    | : :|||||| | |||||
Qy      25 CFWDNGHLYREDQTSPAPGLRCLNWLDAQS-GLASAPVSGAG-NHSYCRNPDEDPRGPWC 82

Db     168 Y-TVDPNVRH-QSCGIKKCEDA 187
           | : :: |   ::|   :| ::
Qy      83 YVSGEAGVPEKRPCEDLRCPET 104


RESULT    4
ID   P70521      PRELIMINARY;      PRT;   716 AA.
AC   P70521;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   HEPATOCYTE GROWTH FACTOR-LIKE PROTEIN PRECURSOR
DE   (MACROPHAGE STIMULATORY PROTEIN) (MSP).
GN   MSP.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 97011126.
RA   OHSHIRO K., IWAMA A., MATSUNO K., EZAKI T., SAKAMOTO O., HAMAGUCHI I.,
RA   TAKASU N., SUDA T.;
RT   "Molecular cloning of rat macrophage-stimulating protein and its
RT   involvement in the male reproductive system.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 227:273-280(1996).
CC   -!- FUNCTION: PROBABLY HAS NO PROTEOLYTIC ACTIVITY, SINCE CRUCIAL AA
CC       CHARACTERISIC OF SERINE PROTEASES CATALYTIC SITES ARE NOT
CC       CONSERVED.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
DR   EMBL; X95096; E223604; -.
DR   PFAM; PF00051; kringle; 4.
DR   PFAM; PF00089; trypsin; 1.
KW   SIGNAL.
FT   SIGNAL        1     31       POTENTIAL.
FT   CHAIN        32    716       MACROPHAGE STIMULATING PROTEIN.
SQ   SEQUENCE   716 AA;  80733 MW;  7C8133F6 CRC32;

  Query Match           9.5%;  Score 178;  DB 11;  Length 716;
  Best Local Similarity  41.3%;  Pred. No. 2.09e-15;
  Matches    26;  Conservative    9;  Mismatches  26;  Indels   2;  Gaps   2;

Db     110 CIMDNGASYRGTVARTADGLPCQAWSRRFPNDHKYTPTPKNGLEENFCRNPDGDPRGPWC 169
           |: |||  ||   :  | || |  |     :   :| :  | : ::||||| |||||||
Qy      25 CFWDNGHLYREDQTSPAPGLRCLNWLDA-QSGLASAPVSGAG-NHSYCRNPDEDPRGPWC 82

Db     170 YTT 172
           | :
Qy      83 YVS 85


RESULT    5
ID   Q24488      PRELIMINARY;      PRT;   685 AA.
AC   Q24488;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
```

```
DE   NEUROTROPHIC RECEPTOR PRECURSOR.
GN   DROR.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON S; TISSUE=BRAIN;
RX   MEDLINE; 93348222.
RA   WILSON C., GOBERDHAN D.C., STELLER H.;
RT   "Dror, a potential neurotrophic receptor gene, encodes a Drosophila
RT   homolog of the vertebrate Ror family of Trk-related receptor tyrosine
RT   kinases.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 90:7109-7113(1993).
DR   EMBL; L20297; G348104; -.
DR   FLYBASE; FBgn0010407; Ror.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
DR   PFAM; PF00051; kringle; 1.
DR   PFAM; PF00069; pkinase; 1.
KW   SIGNAL.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    685       NEUROTROPHIC RECEPTOR.
SQ   SEQUENCE   685 AA;  78142 MW;  22998F56 CRC32;

  Query Match             9.4%;  Score 177;  DB 5;  Length 685;
  Best Local Similarity  27.4%;  Pred. No. 3.28e-15;
  Matches    26;  Conservative   20;  Mismatches  46;  Indels   3;  Gaps   3;

Db    234 TENCYWEDGSTYRGVANVSASGKPCLRWSWLMKEISDFP-ELIG-QNYCRNPGSVENSPW 291
          :  |:|::|  ||     :|:|  || |      ::  |    | ::|||||     :||
Qy     22 SGGCFWDNGHLYREDQTSPAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRGPW 81

Db    292 CFVDSSRE-RIIELCDIPKCADKIWIAIVGTTAAI 325
          |:| :         |:  :|::    |: : |: |
Qy     82 CYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEI 116


RESULT    6
ID   P70006      PRELIMINARY;      PRT;   717 AA.
AC   P70006;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   HEPATOCYTE GROWTH FACTOR-LIKE PROTEIN PRECURSOR.
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KIDNEY;
RX   MEDLINE; 96404125.
RA   ABERGER F., SCHMIDT G., RICHTER K.;
RT   "The Xenopus homologue of hepatocyte growth factor-like protein is
RT   specifically expressed in the presumptive neural plate during
RT   gastrulation.";
RL   MECH. DEV. 54:23-37(1996).
DR   EMBL; Y08734; E274395; -.
DR   PFAM; PF00051; kringle; 4.
DR   PFAM; PF00089; trypsin; 1.
KW   SIGNAL.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29    717       HEPATOCYTE GROWTH FACTOR-LIKE PROTEIN.
SQ   SEQUENCE   717 AA;  82017 MW;  5E6861D5 CRC32;

  Query Match             9.1%;  Score 172;  DB 13;  Length 717;
  Best Local Similarity  41.9%;  Pred. No. 3.10e-14;
  Matches    26;  Conservative   10;  Mismatches  23;  Indels   3;  Gaps   3;

Db    378 CYHGNGELYRGRVSKTRKGIMCRRWEEKQNDLELSLAQPYLVPLEENYCRNPDRDSHGPW 437
          |:  || |||   :    |: |  | : |: |  | | |   : : :|||||| |::|||

Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLDAQSGLA-S-A-PVSGAGNHSYCRNPDEDPRGPW 81

Db    438 CY 439
          ||
Qy     82 CY 83


RESULT    7
ID   Q64007      PRELIMINARY;      PRT;   728 AA.
AC   Q64007;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   HEPATOCYTE GROWTH FACTOR.
GN   HGF.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94363381.
RA   LEE C.C., KOZAK C.A., YAMADA K.M.;
RT   "Structure, genetic mapping, and expression of the mouse Hgf/scatter
RT   factor gene.";
RL   CELL ADHES. COMMUN. 1:101-111(1993).
DR   EMBL; S71816; G632774; -.
DR   MGD; MGI:96079; HGF.
DR   PFAM; PF00051; kringle; 4.
DR   PFAM; PF00089; trypsin; 1.
SQ   SEQUENCE   728 AA;  82972 MW;  E5876030 CRC32;

  Query Match             9.0%;  Score 169;  DB 11;  Length 728;
  Best Local Similarity  33.7%;  Pred. No. 1.18e-13;
  Matches    32;  Conservative   22;  Mismatches  33;  Indels   8;  Gaps   6;

Db    387 SSGQDCYRGNGKNYMGNLSKTRSGLTCSMWDKNMEDLHRHIFWEPDASKLNKNYCRNPDD 446
          :||  |:: ||  |  : :   :|| |  |   ::  :  :   | ::  | :||||||:
Qy     21 GSG-GCFWDNGHLYREDQTSPAPGLRCLNW---LDA-QSGLASAPVSGAGNHSYCRNPDE 75

Db    447 DAHGPWCY-TGNPLIP-WDYCPISRCEGDTTPTIL 479
          |::||||| :|:: :|    |   ||  :||:  |
Qy     76 DPRGPWCYVSGEAGVPEKRPCEDLRCP-ETTSQAL 109


RESULT    8
ID   Q61662      PRELIMINARY;      PRT;   748 AA.
AC   Q61662;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   HEPATOCYTE GROWTH FACTOR (FRAGMENT).
GN   HGF.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B6SJLF1/J; TISSUE=LIVER;
RX   MEDLINE; 94060105.
RA   LIU Y., MICHALOPOULOS G.K., ZARNEGAR R.;
RT   "Molecular cloning and characterization of cDNA encoding mouse
RT   hepatocyte growth factor.";
RL   BIOCHIM. BIOPHYS. ACTA 1216:299-303(1993).
DR   EMBL; X72307; G433431; -.
DR   MGD; MGI:96079; HGF.
DR   PFAM; PF00051; kringle; 4.
DR   PFAM; PF00089; trypsin; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   748 AA;  85200 MW;  24AE0820 CRC32;

  Query Match             8.9%;  Score 168;  DB 11;  Length 748;
  Best Local Similarity  31.9%;  Pred. No. 1.84e-13;
  Matches    30;  Conservative   23;  Mismatches  34;  Indels   7;  Gaps   5;
```

```
Db    407 SSGQDCYRGNGKNYMGNLSKTRSGLTCSMWDKNMEDLHRHIFWEPDASKLNKNYCRNPDD 466
          :||  |:: ||  |  : :   :|| |  |   ::  :  :   | ::  | :||||||:
Qy     21 GSG-GCFWDNGHLYREDQTSPAPGLRCLNW---LDA-QSGLASAPVSGAGNHSYCRNPDE 75

Db    467 DAHGPWCY-TGNPLIP-WDYCPISRCEGDTTPTI 498
          |::|||||  :|:: :|    |   ||   |: ::
Qy     76 DPRGPWCYVSGEAGVPEKRPCEDLRCPETTSQAL 109


RESULT    9
ID   O18783      PRELIMINARY;      PRT;   806 AA.
AC   O18783;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   PLASMINOGEN.
OS   MACROPUS EUGENII (TAMMAR WALLABY).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; METATHERIA;
OC   DIPROTODONTIA; MACROPODIDAE; MACROPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   LAWN R.M., SCHWARTZ K., PATTHY L.;
RL   SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF012297; G2305256; -.
DR   PFAM; PF00051; kringle; 5.
DR   PFAM; PF00089; trypsin; 1.
SQ   SEQUENCE   806 AA;  90980 MW;  4371C1C6 CRC32;

  Query Match             8.9%;  Score 168;  DB 6;  Length 806;
  Best Local Similarity  35.8%;  Pred. No. 1.84e-13;
  Matches    24;  Conservative   17;  Mismatches  22;  Indels   4;  Gaps   3;

Db    101 SDCKSGNGRNYRGTLSKTKSGITCQKWSDLSPHVPNYAPSKYPDAGLEKNYCRNPDDDVK 160
          : |   ||: ||   :   :|: | :|  | :: :  |::  : || : :||||||:| :
Qy     23 GGCFWDNGHLYREDQTSPAPGLRCLNW--LDAQ-SGLASAPVSGAG-NHSYCRNPDEDPR 78

Db    161 GPWCYTT 167
          |||||  :
Qy     79 GPWCYVS 85


RESULT   10
ID   O35727      PRELIMINARY;      PRT;   597 AA.
AC   O35727;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   FACTOR XII.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   SCHLOESSER M., SCHWAGER S., ENGEL W.;
RL   SUBMITTED (JUL-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; X99571; E277119; -.
DR   PROSITE; PS01253; FIBRONECTIN_1; 1.
DR   PFAM; PF00008; EGF; 2.
DR   PFAM; PF00039; fn1; 1.
DR   PFAM; PF00040; fn2; 1.
DR   PFAM; PF00051; kringle; 1.
DR   PFAM; PF00089; trypsin; 1.
SQ   SEQUENCE   597 AA;  65638 MW;  1EA8CD44 CRC32;

  Query Match             8.8%;  Score 165;  DB 11;  Length 597;
  Best Local Similarity  36.8%;  Pred. No. 6.95e-13;
  Matches    25;  Conservative   16;  Mismatches  21;  Indels   6;  Gaps   6;

Db    217 CYEGRGLSYRGQAGTTQSGAPCQRWTVEATYRNMTEKQALSWGLGHHAFCRNPDNDTR-P 275
          |:   |  || :  :  :|  |  | ::|    ::   ::| | |:|::|||||:| | |
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW-LDAQ-SGLAS-APVS-GAGNHSYCRNPDEDPRGP 80

Db    276 WCFVWSGD 283
          ||:| ||:
Qy     81 WCYV-SGE 87


RESULT   11
ID   Q14870      PRELIMINARY;      PRT;   711 AA.
AC   Q14870;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MACROPHAGE-STIMULATING PROTEIN PRECURSOR.
GN   MST1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 93340141.
RA   YOSHIMURA T., YUHKI N., WANG M.H., SKEEL A., LEONARD E.J.;
RT   "Cloning, sequencing, and expression of human macrophage stimulating
RT   protein (MSP, MST1) confirms MSP as a member of the family of kringle
RT   proteins and locates the MSP gene on chromosome 3.";
RL   J. BIOL. CHEM. 268:15461-15468(1993).
DR   EMBL; L11924; G398038; -.
DR   PFAM; PF00051; kringle; 4.
DR   PFAM; PF00089; trypsin; 1.
KW   SIGNAL.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    483       MACROPHAGE-STIMULATING PROTEIN.
FT   CHAIN       484    711       MACROPHAGE-STIMULATING PROTEIN.
SQ   SEQUENCE   711 AA;  80353 MW;  1445C010 CRC32;

  Query Match             8.8%;  Score 166;  DB 4;  Length 711;
  Best Local Similarity  32.5%;  Pred. No. 4.47e-13;
  Matches    27;  Conservative   17;  Mismatches  35;  Indels   4;  Gaps   4;

Db    110 CIMNNGVGYRGTMATTVGGLPCQAWSHKFPNDHKYTPTLRNGLEENFCRNPDGDPGGPWC 169
          |: :||  ||   ::   || |  |     :    :|    | : ::||||| || ||||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLDA-QSGLASAPVSGAG-NHSYCRNPDEDPRGPWC 82

Db    170 YTT-DPAVRF-QSCGIKSCREAA 190
          | : :::|   ::|    | |::
Qy     83 YVSGEAGVPEKRPCEDLRCPETT 105


RESULT   12
ID   Q13350      PRELIMINARY;      PRT;   711 AA.
AC   Q13350;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   HEPATOCYTE GROWTH FACTOR-LIKE PROTEIN.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96224125.
RA   WALTZ S.E., GOULD F.K., AIR E.L., MCDOWELL S.A., DEGEN S.J.F.;
RT   "Hepatocyte nuclear factor-4 is responsible for the liver-specific
RT   expression of the gene coding for hepatocyte growth factor-like
RT   protein.";
RL   J. BIOL. CHEM. 271:9024-9032(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92002016.
RA   HAN S., STUART L.A., DEGEN S.J.;
RT   "Characterization of the DNF15S2 locus on human chromosome 3:
RT   identification of a gene coding for four kringle domains with
```

```
RT   homology to hepatocyte growth factor.";
RL   BIOCHEMISTRY 30:9768-9780(1991).
DR   EMBL; U37055; G1311661; -.
DR   PFAM; PF00051; kringle; 4.
DR   PFAM; PF00089; trypsin; 1.
SQ   SEQUENCE   711 AA;  80319 MW;  4B3C7BC2 CRC32;

  Query Match             8.8%;  Score 166;  DB 4;  Length 711;
  Best Local Similarity  32.5%;  Pred. No. 4.47e-13;
  Matches    27;  Conservative   17;  Mismatches  35;  Indels   4;  Gaps   4;

Db    110 CIMNNGVGYRGTMATTVGGLPCQAWSHKFPNDHKYTPTLRNGLEENFCRNPDGDPGGPWC 169
          |: :||  ||   ::   || |  |     :    :|    | : ::||||| || ||||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLDA-QSGLASAPVSGAG-NHSYCRNPDEDPRGPWC 82

Db    170 YTT-DPAVRF-QSCGIKSCREAA 190
          | : :::|   ::|    | |::
Qy     83 YVSGEAGVPEKRPCEDLRCPETT 105


RESULT   13
ID   Q15146      PRELIMINARY;      PRT;   810 AA.
AC   Q15146;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   PLASMINOGEN PRECURSOR.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BROWNE M.J., CHAPMAN C.G., DODD I., CAREY J.E., LAWRENCE G.M.,
RA   MITCHELL D., ROBINSON J.H.;
RL   FIBRINOLYSIS 0:0-0(0).
DR   EMBL; M74220; G190026; -.
DR   PFAM; PF00051; kringle; 5.
DR   PFAM; PF00089; trypsin; 1.
KW   SIGNAL; ZYMOGEN.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    810       PLASMINOGEN.
SQ   SEQUENCE   810 AA;  90554 MW;  7E939182 CRC32;

  Query Match             8.6%;  Score 161;  DB 4;  Length 810;
  Best Local Similarity  21.8%;  Pred. No. 4.02e-12;
  Matches    27;  Conservative   41;  Mismatches  48;  Indels   8;  Gaps   7;

Db    377 CYHGDGQSYRGTSSTTTTGKKCQSWSSMTPHRHQKTPENYPNAGLTMNYCRNPDAD-KGP 435
          |:  :|: ||   :: : | :| :|  : ::    ::   : ||   :||||||| | :||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNW--LDAQSGL-ASAPVSGAG-NHSYCRNPDEDPRGP 80

Db    436 WCFTT-DPSVR-WEYCNLKKCSGTEASVVAPPPVVLLPDVETPSEEDC-MFGNGKGYRGK 492
          ||: : :::|     |:  :|: | :  :::     :    | |: ::  :|: :::  ::
Qy     81 WCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQVFAPANALPAR 140

Db    493 RATT 496
            ::
Qy    141 SEAA 144


RESULT   14
ID   O46507      PRELIMINARY;      PRT;   334 AA.
AC   O46507;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   PLASMINOGEN (FRAGMENT).
GN   BABPEPSG.
OS   PAPIO HAMADRYAS (HAMADRYAS BABOON).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; PAPIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   COX L.A., JETT C., HIXSON J.E.;
RL   SUBMITTED (OCT-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF029692; G2815618; -.
FT   NON_TER       1      1
SQ   SEQUENCE   334 AA;  36791 MW;  B29A3EA3 CRC32;

  Query Match             8.3%;  Score 156;  DB 6;  Length 334;
  Best Local Similarity  37.9%;  Pred. No. 3.54e-11;
  Matches    25;  Conservative   11;  Mismatches  25;  Indels   5;  Gaps   4;

Db      5 CMFGNGKRYRGKKATTVTGTPCQEWAAKEPHSHLIFTPETYPRAGLEKNYCRNPDGDVGG 64
          |   ||  ||   ::   |  | :|   :::| |  :| :   || : :||||| |  |
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWL--DAQSGLASAPVS-G-AG-NHSYCRNPDEDPRG 79

Db     65 PWCYTT 70
          |||| :
Qy     80 PWCYVS 85


RESULT   15
ID   Q91691      PRELIMINARY;      PRT;   716 AA.
AC   Q91691;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   GROWTH FACTOR LIVERTINE.
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   I ALTABA A., THERY C.;
RL   SUBMITTED (MAY-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U57455; G1399751; -.
DR   PFAM; PF00051; kringle; 4.
DR   PFAM; PF00089; trypsin; 1.
SQ   SEQUENCE   716 AA;  81971 MW;  14BE3913 CRC32;

  Query Match             8.3%;  Score 157;  DB 13;  Length 716;
  Best Local Similarity  37.1%;  Pred. No. 2.29e-11;
  Matches    23;  Conservative   12;  Mismatches  24;  Indels   3;  Gaps   3;

Db    377 CYHGNGELYSGRVSKTRKGIKCRRWEEKRNDLELSLDQPYLVPLEENYCRNPDRDSHGPW 436
          |:  || ||    :    |::|  | : :: |  |   |   : : :|||||| |::|||
Qy     25 CFWDNGHLYREDQTSPAPGLRCLNWLDAQSGLA-S-A-PVSGAGNHSYCRNPDEDPRGPW 81

Db    437 CY 438
          ||
Qy     82 CY 83


Search completed: Fri Sep 17 18:28:51 1999
Job time : 74 secs.
```